haematologica
Journal of the Ferrata Storti Foundation

# Gut microbiome alterations at acute myeloid leukemia diagnosis are associated with muscle weakness and anorexia

by Sarah A. Pötgens, Violaine Havelange, Sophie Lecop, Fuyong Li, Audrey M. Neyrinck, Florence Bindels, Nathalie Neveux, Jean-Baptiste Demoulin, Ine Moors, Tessa Kerre, Johan Maertens, Jens Walter, Hélène Schoemans, Nathalie M. Delzenne, and Laure B. Bindels

*Received: August 25, 2023.*
*Accepted: March 19, 2024.*

*Citation: Sarah A. Pötgens, Violaine Havelange, Sophie Lecop, Fuyong Li, Audrey M. Neyrinck, Florence Bindels, Nathalie Neveux, Jean-Baptiste Demoulin, Ine Moors, Tessa Kerre, Johan Maertens, Jens Walter, Hélène Schoemans, Nathalie M. Delzenne, and Laure B. Bindels. Gut microbiome alterations at acute myeloid leukemia diagnosis are associated with muscle weakness and anorexia. Haematologica. 2024 Mar 28. doi: 10.3324/haematol.2023.284138 [Epub ahead of print]*

# Gut microbiome alterations at acute myeloid leukemia diagnosis are associated with muscle weakness and anorexia

Sarah A. Pötgens[1], Violaine Havelange[2,3], Sophie Lecop[1], Fuyong Li[4,5], Audrey M. Neyrinck[1], Florence Bindels[6], Nathalie Neveux[7], Jean-Baptiste Demoulin[3], Ine Moors[8], Tessa Kerre[8], Johan Maertens[9,10], Jens Walter[11], Hélène Schoemans[9,12], Nathalie M. Delzenne[1], Laure B. Bindels[1,13]

[1]Metabolism and Nutrition Research Group, Louvain Drug Research Institute, UCLouvain, Université catholique de Louvain, Brussels, Belgium.

[2]Department of Hematology, Cliniques Universitaires Saint-Luc, UCLouvain, Université catholique de Louvain, Brussels, Belgium.

[3]Experimental Medicine Unit, De Duve Institute, UCLouvain, Université catholique de Louvain, Brussels, Belgium.

[4]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Kowloon, Hong Kong SAR, China.

[5]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada.

[6]Maison Médicale de Grez-Doiceau, Grez-Doiceau, Belgium.

[7]Clinical Chemistry Department, Cochin Hospital, Paris Centre University Hospitals, Paris, France

[8]Department of Hematology, Ghent University Hospital, Ghent University, Ghent, Belgium.

[9]Department of Hematology, University Hospitals Leuven, Leuven, Belgium.

[10]Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium

[11]Department of Medicine, School of Microbiology, APC Microbiome Ireland, University College Cork, Cork, Ireland.

[12]Department of Public Health and Primary Care, ACCENT VV, KU Leuven, Leuven, Belgium

[13]Welbio Department, WEL Research Institute, Wavre, Belgium

**Authors' contributions**

Conception and design of the study: NMD and LBB. Supervision of the work: LBB. Contributions to the design of the clinical study: VH, JBD, HS. Clinical data collection and biological sampling: SAP, VH, FB, IM, TK, JM, HS, and LBB. Clinical data analysis: SAP. Metabolomics analysis: SAP with the help of SL for urinary metabolomics. Microbiome analyses: SAP, FL, JW, and LBB. Analyses of metabolic and inflammatory markers SAP, ANM, and LBB. Citrulline analysis by mass spectrometry: NN. Data integration: SAP and LBB. Data interpretation: SAP and LBB. Contribution to data interpretation: VH, HS, and NMD. Acquisition of funding: LBB. Drafting of the article: SAP and LBB. Revision of the article: all authors. Final approval of the version to be published: all.

**Running heads**

Gut microbiota in AML

**Corresponding author**

Prof. Laure B. Bindels, Metabolism and Nutrition Research Group, Louvain Drug Research Institute, UCLouvain, Université catholique de Louvain, UCLouvain, avenue E. Mounier box B1.73.11, B-1200 Brussels, Belgium; E-mail address: laure.bindels@uclouvain.be ORCID: LBB 0000-0003-3747-3234

**Data-sharing statement**

The full details of the methods described in this paper are provided in the Supporting Information. Raw sequences from 16S rRNA gene sequencing and metagenomics can be found in the SRA database (project ID: PRJNA813705).

**Word count**

Abstract: 251

Main text: 3946

Number of figures: 7

Number of tables: 1

Number of supplementary files: 1

## Clinical study registration

Retrospectively registered on March 20, 2018 on Clinicaltrials.gov: NCT03881826.


## Acknowledgments

The authors are grateful to all the study participants. The authors are grateful to Myriam Cleeren, Jill Pannecoucke, Vera Ciku, Dr. Martina Sboarina, Dr. Morgane Thibaut, Isabelle Blave and Dr. Julie Vanacker as well as the data nurses and the medical team involved in the study, for their help in the collection of biological samples and clinical data. SAP and LBB thank Martin Nicolas and Bouazza Es Saadi for providing technical assistance during sample analyses and Pascaline Incourt for her help with the analyses of the food questionnaires. SAP and LBB thank Dr. Nicolas Joudiou and the UCLouvain-LDRI Nuclear and Electron Spin Technologies (NEST) platform for providing easy access and skilled assistance to the NMR spectrometer and Prof. Ana Beloqui for access to a Meso Scale Discovery microplate reader. The authors thank Dr. Inge Huybrechts and Dr. Ernst Rietzschel for sharing the food frequency questionnaires. The authors also acknowledge the Centre d'expertise et de services Génome Québec for their sequencing services and support (shotgun metagenomics), and the University of Minnesota Genomics Centre for (16S rRNA gene sequencing). This research has benefitted from the statistical consult with Statistical Methodology and Computing Service, a technological platform at UCLouvain – SMCS/LIDAM, UCLouvain.

## Funding

This work was funded by an ESPEN Research fellowship awarded to LBB, the Fonds de la Recherche Scientifique – FNRS (F.R.S.-FNRS) under Grant MIS F.4512.20, the Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO), and the F.R.S. – FNRS under EOS Project No. 40007505. This work would not have been possible without the support of the Télévie (Intercachectomics Consortium) and the Louvain Foundation. LBB is a Collen-Francqui Research Professor (Francqui Foundation) and the recipient of subsidies from the FSR (Fonds Spéciaux de la Recherche, UCLouvain, including the Action de Recherche Concertée LIPOCAN (19-24.096)) and from the Walloon Region in the context of the funding of the strategic axis FRFS-WELBIO (40009849). NMD is a recipient of grants from the Fonds de la Recherche Scientifique (FRS-FNRS) (PINT-MULTI R.8013.19 [NEURON-ERANET, call 2019] and PDR

T.0068.19). The funders had no role in the study design, data collection and analysis, interpretation of results, decision to publish, or preparation of the manuscript.

**Competing interests**

HS reports having received personal fees from Incyte, Janssen, Novartis, Sanofi and from the Belgian Hematological Society (BHS), as well as research grants from Novartis and the BHS, all paid to her institution. She has also received non-financial support (travel grants) from Gilead, Pfizer, the EBMT (European Society for Blood and Marrow transplantation) and the CIBMTR (Center for International Bone Marrow Transplantation Research). None of the potential conflicts of interest are relevant to this project. All other authors declare no conflicts of interest.

**Abstract**

The gut microbiota makes critical contributions to host homeostasis, and its role in the treatment of acute myeloid leukaemia (AML) has attracted attention. We investigated whether the gut microbiome is affected by AML, and whether such changes are associated with cachectic hallmarks. Biological samples and clinical data were collected from 30 antibiotic-free AML patients at diagnosis and matched volunteers (1:1) in a multicenter cross-sectional prospective study. The composition and functional potential of the faecal microbiota were analyzed using shotgun metagenomics. Faecal, blood, and urine metabolomics analyses were performed.

AML patients displayed muscle weakness, anorexia, signs of altered gut function, and glycaemic disorders. The composition of the faecal microbiota differed between patients with AML and control subjects, with an increase in oral bacteria. Alterations in bacterial functions and faecal metabolome support an altered redox status in the gut microbiota, which may contribute to the altered redox status observed in patients with AML. *Eubacterium eligens*, reduced 3-fold in AML patients, was strongly correlated with muscle strength and citrulline, a marker of enterocyte mass and function. *Blautia* and *Parabacteroides*, increased in patients with AML, were correlated with anorexia. Several bacterial taxa and metabolites (*e.g. Blautia*, *Prevotella*, phenylacetate, and hippurate) previously associated with glycaemic disorders were altered. Our work revealed important perturbations in the gut microbiome of AML patients at diagnosis, which are associated with muscle strength, altered redox status, and anorexia. These findings pave the way for future mechanistic work to explore the function and therapeutic potential of the bacteria identified in this study.

Keywords: MicroAML study, cachexia, GDF15, FGF21, *Lachnospira eligens*, HOMA-IR2.

## Introduction

The role of the gut microbiota (GM) in acute myeloid leukaemia (AML) has attracted the attention of the scientific community. The complex microbial community inhabiting the human gastrointestinal tract cultivates an intricate and mutually beneficial relationship with its host. The GM synthesizes essential metabolites and releases bioactive compounds, which can influence host physiology[1]. 15% of the metabolites found in mammalian blood are derived from the GM[2], underlying the deep imbalance between the host and GM metabolism. The GM can also regulate host metabolism and immunity through the regulation of gut function[3]. Therefore, GM is considered a key regulator of host metabolism and inflammatory status in different pathophysiological contexts, including metabolic syndrome and insulin resistance[4].

AML is a clonal disorder of hematopoietic stem cells, resulting in impaired production of the myeloid blood cell lineage. The standard treatment for younger AML patients (< 65 years) is intensive induction chemotherapy followed by consolidation chemotherapy, and in some cases, allogeneic hematopoietic stem cell transplantation (HSCT). The GM seems to control aspects of haematopoiesis through mechanisms including metabolite production as well as through circulating microbial DNA[5]. AML is also characterized by an altered host immunity, the latter being widely influenced by the GM. Despite the expected links between AML and GM, little is known about the influence of AML on GM composition and more is known about the impact of chemotherapy.

Indeed, patients undergoing induction chemotherapy experience loss of microbiota diversity[6, 7]. The reduction in some beneficial bacterial species, such as *Faecalibacterium prausnitzii* and *Bifidobacterium* spp., persisted even after completion of chemotherapy[8]. A higher microbial diversity at diagnosis was linked to a reduced risk of infectious complications following induction chemotherapy, suggesting an interest in microbiota profiling for infectious risk stratification[9]. Much of the recent work investigating the GM in the context of AML has focused on its relationship with the development of graft-*versus*-host disease (GVHD) after allogeneic HSCT. Higher α-diversity was associated with better survival[10] and higher levels of *Blautia* with reduced GVHD mortality[11]. The presence of *Eubacterium limosum* is also associated with a reduced risk of disease relapse[12]. During the post-transplant period, the presence of antibiotic-resistant bacteria or domination by some bacteria, such as *Enterococcus* species and

*Proteobacteria*, is associated with a higher risk of infection[13, 14]. These observations have encouraged the evaluation of strategies to enhance microbial diversity. In this context, a recent study investigated the safety and diversity-enhancing ability of autologous faecal microbiota transfer in patients with AML receiving intensive chemotherapy and antibiotics[15]. Faecal material collected at the time of diagnosis was used for faecal microbiota transfer. This transfer appeared to be safe and could restore microbial richness and diversity based on α-diversity indices.

Many unknowns remain concerning the composition and activity of GM at diagnosis and its potential influence on the host in AML. In a mouse model of leukaemia, we showed that GM composition was altered alongside the gut barrier and that reversing these changes through pre- and probiotics improved cachectic features[16, 17]. Cachexia is a multifactorial systemic syndrome characterized by substantial weight loss (coming from atrophy of the skeletal muscle and adipose tissue), often accompanied by anorexia, muscle weakness, insulin resistance, and inflammation, which occurs in approximately 50–80% of patients with cancer[18]. Although cachexia affects 40% of patients with hematological cancer and is often witnessed in clinics, studies focusing on AML-related cachexia are scarce[19]. Therefore, we launched an academic multicentric cross-sectional prospective study to evaluate GM composition and activity in AML patients and volunteers (CT) matched for age, sex, BMI and smoking status. Thirty patients newly diagnosed with AML were recruited before therapeutic intervention. We collected information on the patients' food habits and cachectic hallmarks. We also assessed the inflammatory status of the patients, as well as metabolic and gut function markers. Through a multi –*omics* integrative approach combining metagenomics, faecal, blood, and urine metabolomics, and clinical data, we investigated the GM of AML patients at diagnosis and their links to cachectic hallmarks.

## Methods

### *Subjects*

Thirty patients newly diagnosed with AML were recruited between December 2015 and December 2019 from three Belgian University hospitals (Saint-Luc Brussels (n = 13), UZ Leuven (n = 15), and UZ Gent (n = 2)). The exclusion criteria included antibiotic consumption within the last 30 days, chronic intestinal diseases, obesity (body mass index (BMI) > 30), pregnancy, gastric bypass, and treatment with antidiabetic or hypoglycaemic drugs.

Control (CT) subjects from the general population were recruited between December 2017 and January 2020 based on the same inclusion/exclusion criteria, except for the AML diagnosis. They were matched (1:1) for age, sex, BMI, and smoking status, all factors known to impact GM. This study followed the ethical guidelines set out in the Declaration of Helsinki, was approved by the "Comité d'éthique Hospitalo-facultaire des Cliniques Universitaires Saint-Luc" (B403201317128), and all participants provided written informed consent. A full description of the study design is provided in the Supporting Information.

### *Sample and data collection*

All biological sampling and data collection were performed at diagnosis (*i.e.* before the initiation of chemotherapy and antibiotic treatment). The full details are provided in the Supporting Information.

### *Sample and data analyses*

Biochemical, gut microbiota, $^{1}$H-NMR and statistical analyses are detailed in the Supporting Information. Statistical significance was set at $p < 0.05$ and q-value < 0.1.

## Results

### *CT and AML patients display similar anthropometric characteristics and habits*

As ensured by one-to-one matching, the CT and AML patient groups did not display any differences in terms of age, sex, BMI, body composition, and smoking percentage (Table 1). They also had similar habits in terms of daily alcohol consumption. The food frequency questionnaire[20] revealed no differences in the overall dietary score. No differences between the groups were detected for drug exposure, food supplement consumption, and transit scores (Table S1, Figure S1).

### *AML patients display specific GM alterations*

GM composition was assessed in the same stool samples using two independent techniques (16S rRNA gene sequencing and metagenomics). GM composition was not differentially clustered between CT and AML patients, as revealed by PCA at the genus level (Figure 1A, Figure S2A).

Permutational multivariate analysis of variance (PERMANOVA) revealed that 3.1% of the variation in GM composition, as assessed at the genus level by shotgun metagenomics, was explained by AML (Figure 1A-B). At the species level, the disease explained 3.2% of the variation of the dataset (Figure S3A). We also assessed the explanatory power of other factors previously shown to explain most of the variation in GM composition in the Belgian Flemish Gut Flora Project, a cohort of 1106 Belgian volunteers[21], namely BMI, hemoglobin, and age (Figure 1B and Figure S3B). Disease presence had the highest explanatory power, close to BMI (3.1% in this cohort and ~3% in the Belgian Flemish Gut Flora Project) while age and sex were not significant. Total bacterial levels were similar between both groups (Figure S2B), as were the α-diversity indices (Figure S2C). However, the univariate analyses revealed specific changes. Three genera, namely *Actinomyces*, *Blautia* and P*arabacteroides*, and one species, *Parvimonas micra*, were increased in the faeces of patients with AML, whereas *Eubacterium eligens* was decreased (Figure 1C). These changes, highlighted by shotgun metagenomics, were confirmed by 16S rRNA gene sequencing, with the exception of *Blautia* (Figure S2D, Table S2). *Coprococcus* and *Prevotella* were reduced in AML patients using 16S rRNA gene sequencing ($q = 0.109$ using shotgun metagenomics) (Figure S2D, Table S2). The relative abundances of oral species increased more than two-fold, while obligate relative abundances of anaerobe genera were reduced (Figure 1D).

PCA of metagenomic functions (Figure 1E) confirmed the compositional analyses, with no clear global functional changes between the two groups. However, 22 functions were significantly different (Figure 1F, Table S3). Of these 22 functions, only four were decreased in AML patients, whereas the others were increased. The contributions of the species and genera to each of these functions are shown in Figure S4. A visual inspection of such graphs indicates that for the majority of the functions, the changes in their functional abundances observed in AML patients cannot be ascribed to specific species/genera, thereby suggesting that functional changes in the GM may be independent of compositional changes. These compositional and functional changes were sufficient to predict disease status in a testing set with an accuracy of 87% based on the top altered bacteria and 77% based on the altered functions using Random Forest models (Figure S5).

### *AML patients demonstrate anorexia, muscle weakness and glycaemic disorders*

Patients with AML displayed changes in hemoglobin levels and white blood cell count, C-reactive protein (CRP), albumin, and glycaemia levels characteristics of the disease (Figure 2, Figure S6). They also had higher CRP and lower albumin levels, leading to a higher modified Glasgow prognostic score (mGPS) (Figure 2A). In addition to reflecting a higher inflammatory status, mGPS has been recently associated with adverse outcomes in patients with newly diagnosed AML[22]. Appetite score and muscle strength were reduced in patients with AML (Figure 2B). Fasted glycaemia and insulinemia, as well as HOMA-IR2, were higher in AML patients (Figure 2C), collectively reflecting and confirming glucose metabolism alterations in AML[23]. This appears especially the case for a subset of patients that, intriguingly, presented a specific microbial signature (Figures S7, S8 and supplementary results and discussion).

In addition to an increase in CRP, other inflammatory markers (*i.e.* interleukin-6 (IL6), interleukin-8 (IL8), monocyte chemoattractant protein 1 (MCP1), and tumor necrosis factor alpha (TNFα)) were elevated in AML patients as compared to CT (Figure 3A). Together with the increase in interleukin-10 (IL10) and reduction in transforming growth factor beta-1 (TGFβ1), this confirms alterations in cytokine levels in AML patients[24] and a high inflammatory status. We also reported an increase in growth differentiation factor-15 (GDF15), a member of the TGFβ superfamily, and fibroblast growth factor 21 (FGF21) in AML patients (Figure 3B). GDF15 mediates tumor-induced anorexia and body weight loss[25] while FGF21 has been implicated in fasting-induced muscle atrophy and weakness in mice[26].

Plasma levels of citrulline, a non-proteinogenic amino acid reflecting enterocyte mass and function[27], were decreased in patients with AML (Figure 3C). Together with the increased level of lipopolysaccharide-binding protein (LBP) (Figure 3C), the reduction in citrulline levels in AML patients suggests an alteration of the gut function.

### *AML patients display specific faecal, blood and urine metabolome alterations*

PCA performed on faecal, blood, and urine metabolites did not reveal a clear clustering between CT and AML (Figure S9), with AML samples being more dispersed than the CT samples. PERMANOVA revealed that 11.2% and 2.9% of the variation in blood and urine metabolomics, respectively, can be explained by the presence of AML ($p < 0.05$). The metabolites driving the

separation between both groups were identified using multivariate discriminant analysis (Figure S9). Univariate analysis also revealed significantly altered metabolites (Figure 4, Table S4). In AML faeces, two sugars (glucose and galactose), three alcohols (methanol, ethanol, and glycerol) and two amino acids (phenylalanine and threonine) decreased. Bacterial amino acid-derived metabolites were modified: 3-phenylpropionate decreased, while phenylacetate increased. In AML blood, two sugars (glucose and mannose), as well as ethanol, increased. Amino acids (alanine, asparagine, citrulline, cystine, glutamine, methionine, sarcosine, and threonine) decreased, except for glutamate, which increased. Uridine and urea levels were decreased. Lactate levels increased, while acetate, fumarate, and succinate levels decreased. Amino acid catabolic products, such as 3-hydroxyisovalerate, 2-hydroxybutyrate, formate, and dimethylamine, were increased. Ascorbate, also known as vitamin C, was decreased in AML patients. In the urine of AML patients, the metabolites were mainly increased. This was the case for most amino acids (2-aminobutyrate, 3-aminoisobutyrate, carnitine, isoleucine, leucine, taurine, threonine, and tyrosine) and nucleic base-related metabolites (hypoxanthine and pseudouridine), with the exception of uracil, which decreased. Formate, maleate, and Sumiki's acid also increased, whereas hippurate and trigonelline decreased. Consistent with the alterations in glucose metabolism mentioned earlier, glucose levels were also increased in the AML urine samples.

### *Altered bacteria and functions in AML patients correlate with several faecal, blood and urine metabolites*

Next, we performed correlations between the top altered bacterial taxa/functions and faecal, blood, and urine metabolite levels. The genus *Eubacterium*, as well as the associated species *E. eligens* and *E. hallii*, showed the strongest correlations with faecal metabolites, especially malonate (negative correlation) and glycerol (positive correlation) (Figure 5). Several blood metabolites were strongly correlated with the different bacterial taxa (Figure 5). Urea was negatively correlated with *Blautia* genus and positively correlated with the *Sutterellaceae* family and *E. eligens.* Ascorbate levels were negatively correlated with *Clostridium spiroforme*. 3-hydroxyisobutyrate correlated negatively with the *Erysipelatoclostridium* genus, and 2-hydroxybutyrate correlated positively with the *Parabacteroides* genus. The correlations were weaker for urine metabolites (Figure S10). *E. eligens* displayed the highest number of

correlations (all negative), mainly with amino acids. *Parabacteroides* also showed several positive correlations. Most of these correlations were maintained when corrected for age (Figure S11). Among those correlations, faecal malonate and glycerol correlations with *Eubacterium*, *E. eligens*, and *E. hallii* were also found in the healthy control group alone (Figure S12). In the healthy control group, blood urea also correlated with *Blautia* as well as *E. eligens*, and 3-hydroxyisobutyrate correlated with *Erysipelatoclostridium* genus (Figures S13, S14).

Interestingly, when looking at the correlations between altered microbial functions and faecal, blood, and urine metabolomes, some faecal and blood metabolites demonstrated strong correlations with many different functions (Figure S15). For instance, in faeces, glycerol showed strong negative correlations with almost all increased functions in the AML group. In blood, several metabolites showed multiple correlations with bacterial function. Some were positive (*e.g.* correlation with formate) and some were negative (*e.g.* correlation with ascorbate and propylene glycol). In contrast, ferredoxin-nitrite reductase was negatively correlated with many urine metabolites, whereas glycerol dehydratase was positively correlated with similar metabolites (Figure S16).

### *Altered bacteria and functions in AML patients associate with clinical, metabolic and inflammatory parameters*

We then looked for associations between altered bacterial taxa and functions and the collected clinical, dietary, inflammatory, and metabolic parameters in the whole cohort and exclusively in the AML group to delineate the strongest correlations. We found that many taxa, including *Parvimonas micra*, *Actinomyces*, *Parabacteroides* and *Blautia*, were positively associated with inflammatory markers in the entire cohort (Figure 6). *E. eligens*, decreased in AML patients, was negatively associated with inflammation and positively associated with citrulline and muscle strength, while the latter correlation was maintained inside the AML group (Figure S17). Consistent with these findings, muscle strength was strongly correlated with blood citrulline (rho = 0.3522, p = 0.0058) and CRP levels (rho = -0.4585, p = 0.0002). Looking at the negative correlations, we noticed that *Blautia* was inversely associated with muscle strength. Some parameters, such as haemoglobin, IL6 and IL8 as well as TGFβ1, and citrulline, were associated with many different functions (Figure S18), with some associations maintained when considering

exclusively the AML patients (Figure S19). Most of these correlations were maintained when corrected for age (Figure S20).

**Discussion**

This study highlights alterations in GM composition and activity, cachectic hallmarks and metabolic and inflammatory disturbances in patients newly diagnosed with AML, with several associations found between these alterations (Figure 7).

Altered bacterial functions in AML patients are genes catalyzing redox reactions, suggesting an altered redox status in the GM of AML. Two elements from faecal metabolomics further support the hypothesis of redox imbalance in AML GM towards a more pro-oxidative status. First, phenylalanine, which is decreased in the faeces of patients with AML, can be metabolized by the GM into 3-phenylpropionate and phenylacetate. 3-phenylpropionate, reduced in AML patients, is produced through a reductive pathway, while phenylacetate, which is increased in AML patients, arises through an oxidative pathway[28]. Along these lines, faecal phenylacetate was associated with a higher pro-oxidant and pro-inflammatory status in elderly volunteers[29] whereas blood phenylacetate was elevated in cancer cachectic patients[30]. Second, glycerol levels were reduced in AML faeces. This decrease could result from an oxidative stress on the bacterial membranes, requiring increased lipid synthesis. This is supported by the increase in lipid synthesis functions (glycerol dehydratase and lipoate protein ligase), as well as by the correlations between glycerol and redox and lipid synthesis functions. Notably, the reduction in the relative abundance of obligate anaerobic genera may reflect increased oxygen levels in the intestine, contributing to the altered bacterial redox status. Several elements from blood and urine metabolomics indicate the presence of oxidative stress on the host side, which is consistent with previous studies[31]. First, taurine, a non-proteinogenic amino acid with antioxidant properties, was more excreted in the urine of patients with AML. Second, ascorbate, another antioxidant, was decreased in the blood of AML patients compared with CT. Interestingly, a few preclinical studies have demonstrated that the GM affects the host antioxidant defence system[32]. Considering all these elements, we propose that the altered activity and composition of the GM observed in AML patients could contribute to the oxidative stress present in these patients. This hypothesis is in line with the results of a previous study arguing for a microbial role in the oxidative stress present in some

patients with acute leukaemia under chemotherapy. Specifically, the authors proposed oxidative stress as a mediator involved in *Akkermansia*-related neutropenic fever[33].

In addition to their link with the GM redox pathways, some metabolites pointed out previously could be linked to altered glycaemic homeostasis in AML patients. Indeed, an increase in systemic phenylacetate was causally associated with insulin resistance[34] and 2-hydroxybutyrate, a marker for insulin resistance[35], was increased in the blood of our AML patient cohort. Hippurate, a glycine conjugate of microbial benzoate, was decreased in the urine of AML patients. This metabolite is commonly seen as a general marker of metabolic health and is causally associated with glucose tolerance[36]. Hippurate and *Prevotella* increased following the administration of fibre[37]. Their positive correlation in the current study suggests that reduced hippurate levels may be explained by a decrease in *Prevotella*. *Blautia* has been inconsistently associated with different aspects of metabolic syndrome[38]. This genus is increased in anorectic patients[39]. Consistent with this finding, in our study, *Blautia* increased in AML patients and correlated negatively with appetite and positively with the pro-anorexigenic cytokine GDF15. Collectively, our data reinforce previous observations of the contribution of the GM to metabolic derangements in acute leukaemia preclinical models[40] and suggest that the GM could influence glucose metabolism and energy intake in AML patients, strengthening the relevance of GM alterations in this pathological context.

Among all the alterations and correlations with the most altered bacteria, *E. eligens* stands out. This species, reduced three-fold in leukemic patients compared to CT, was strongly correlated with muscle strength in the entire cohort and when considering exclusively AML patients. Based on urinary metabolomics, we found that *E. eligens* was widely associated with the urinary excretion of amino acids and amino acid metabolites. Higher urinary excretion of carnitine and 2-aminobutyrate was previously associated with cachexia and muscle loss[41]. Blood metabolomics revealed a strong association between *E. eligens* and urea. Such associations with urea, a metabolite involved in the regulation of ammonium levels, may be related to the urinary excretion of amino acids, representing a loss of ammonium. The increase in 8-oxoguanine deaminase and caffeine dehydrogenase, which are two functions strongly correlated with urea, could also indicate greater ammonium use by the GM. In elderly individuals, *E. eligens* was positively associated with markers of lower frailty[42] reinforcing the relevance of the association

of this bacterial species with muscle strength. Considering the growing awareness of the importance of muscle wasting in AML outcomes[19], GM modulation could be an innovative strategy to help patients fight muscle weakness.

*E. eligens* was also correlated with the blood levels of citrulline. Reduction in citrulline levels following intensive chemotherapy was previously reported in AML mice and patients[43] and low plasma citrulline levels before allogeneic HSCT were associated with a higher risk of gastrointestinal GVHD and non-relapse mortality in patients[44]. Citrulline decrease combined with an increase in systemic LBP levels suggests an alteration of the gut function in AML patients at diagnosis. Gut permeability was not assessed at the functional level in this study as the procedure for ingestible probes is incompatible with the clinical management of patients at diagnosis. An alteration in intestinal functions, such as absorption and gut barrier, could potentially contribute to the inflammation, metabolic disorders and oxidative stress (the latter through reduced ascorbate absorption) found in these patients.

*Parabacteroides*, especially *P. merdae* and *P. distasonis*, were increased in patients with leukaemia. These two genera were reported to be affected in many pathologies[38] and were ascribed both beneficial and harmful effects. An increase in *Parabacteroides* was also found in leukemic mice and was attributed to a reduction in food intake[17]. Consistently, in our dataset, *P. merdae* negatively correlated with appetite. In mice with acute leukaemia, we found a bloom in members of the *Enterobacteriaceae* family[17]. In our AML cohort, *Enterobacteriaceae* levels did not change (Figure S21). *Prevotella* showed significantly lower abundance in leukemic patients, a reduction mainly explained by a six-fold reduction in *P. copri*, a genus also reduced in cachectic patients[45]. *Actinomyces* and *P. micra* were increased in our cohort of AML patients. These two oral bacteria were previously found to be increased in colorectal cancer patients[38, 46]. In line with these studies, the relative abundance of oral species increased more than twice. Translocation of oral bacteria to the gut has been associated with many severe diseases, such as inflammatory bowel disease and liver cirrhosis and can promote intestinal inflammation[47]. Reciprocally, we hypothesize that modifications of the microbiota composition observed upon AML are triggered by inflammation, a key driver of microbial dysbiosis[48, 49].This hypothesis is coherent with the positive correlations found between inflammatory cytokines and bacteria increased in AML. Based on preclinical studies[50], we further speculate that this dysbiosis

contributes to the disease progression, hence the notion that AML and dysbiosis may be co-evolving.

Our clinical data are consistent with our previous preclinical observations[16, 17]. In both settings, we observed metabolic and inflammatory alterations, impaired gut function, and altered GM composition and function. We previously showed that counteracting microbial alterations using rationally selected pre- and/or probiotics, including fibres, improved gut function, and counteracted muscle atrophy in leukemic mice[16, 17]. These observations further support the therapeutic potential of GM-based adjuvant treatments.

This study has several limitations. The sample size was limited, and the data from this exploratory pilot study will need to be confirmed in a larger cohort. As the primary objective of this study was to evaluate GM composition and function, we enforced several exclusion criteria for metabolic and pathological situations already known to be associated with changes in GM. The exclusion of patients treated with antibiotics within 30 days and obese patients led us to exclude a considerable proportion of the patients referred to the hospital centers involved in the study. Impedancemetry is less precise than CT-scan in evaluating body composition, but CT-scan was not compatible with the clinical management of patients at diagnosis. In addition, the analysis of skeletal muscle function would have benefited from whole-body strength and range of movement tests.

In conclusion, our work reveals important alterations in the GM composition and function of AML patients at diagnosis before any therapeutic intervention. This finding may be of clinical importance considering that autologous FMT is emerging as an option for these patients[15]. Our findings call for caution when using autologous faecal material transfer during the therapeutic care of AML patients and goes in favour of heterologous transfer to increase the gut microbiota diversity and richness in these patients. Whether such FMT would also benefit cachexia will need to be determined. Furthermore, these GM alterations are associated with cachectic hallmarks (muscle weakness, anorexia, and inflammation), redox status, and signs of gut dysfunction. However, association does not imply causality. Therefore, our findings will constitute the basis of future mechanistic studies exploring the contribution and therapeutic potential of the bacteria identified in this study, such as *Eubacterium eligens*.

**References**


1. Postler TS, Ghosh S. Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. Cell Metab. 2017;26(1):110-130.
2. Visconti A, Le Roy CI, Rosa F, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun. 2019;10(1):4505.
3. Tilg H, Zmora N, Adolph TE, Elinav E. The intestinal microbiota fuelling metabolic inflammation. Nat Rev Immunol. 2020;20(1):40-54.
4. Delzenne NM, Knudsen C, Beaumont M, Rodriguez J, Neyrinck AM, Bindels LB. Contribution of the gut microbiota to the regulation of host metabolism and energy balance: a focus on the gut-liver axis. Proc Nutr Soc. 2019;78(3):319-328.
5. Manzo VE, Bhatt AS. The human microbiome in hematopoiesis and hematologic disorders. Blood. 2015;126(3):311-318.
6. Rashidi A, Kaiser T, Graiziger C, et al. Gut dysbiosis during antileukemia chemotherapy versus allogeneic hematopoietic cell transplantation. Cancer. 2020;126(7):1434-1447.
7. D'Angelo CR, Sudakaran S, Callander NS. Clinical effects and applications of the gut microbiome in hematologic malignancies. Cancer. 2021;127(5):679-687.
8. Galloway-Pena JR, Smith DP, Sahasrabhojane P, et al. The role of the gastrointestinal microbiome in infectious complications during induction chemotherapy for acute myeloid leukemia. Cancer. 2016;122(14):2186-2196.
9. Galloway-Pena JR, Shi Y, Peterson CB, et al. Gut Microbiome Signatures Are Predictive of Infectious Risk Following Induction Therapy for Acute Myeloid Leukemia. Clin Infect Dis. 2020;71(1):63-71.
10. Taur Y, Jenq RR, Perales MA, et al. The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. Blood. 2014;124(7):1174-1182.
11. Jenq RR, Taur Y, Devlin SM, et al. Intestinal Blautia Is Associated with Reduced Death from Graft-versus-Host Disease. Biol Blood Marrow Transplant. 2015;21(8):1373-1383.
12. Peled JU, Devlin SM, Staffas A, et al. Intestinal Microbiota and Relapse After Hematopoietic-Cell Transplantation. J Clin Oncol. 2017;35(15):1650-1659.

13. Bilinski J, Robak K, Peric Z, et al. Impact of Gut Colonization by Antibiotic-Resistant Bacteria on the Outcomes of Allogeneic Hematopoietic Stem Cell Transplantation: A Retrospective, Single-Center Study. Biol Blood Marrow Transplant. 2016;22(6):1087-1093.
14. Harris B, Morjaria SM, Littmann ER, et al. Gut Microbiota Predict Pulmonary Infiltrates after Allogeneic Hematopoietic Cell Transplantation. Am J Respir Crit Care Med. 2016;194(4):450-463.
15. Malard F, Vekhoff A, Lapusan S, et al. Gut microbiota diversity after autologous fecal microbiota transfer in acute myeloid leukemia patients. Nat Commun. 2021;12(1):3084.
16. Bindels LB, Beck R, Schakman O, et al. Restoring specific lactobacilli levels decreases inflammation and muscle atrophy markers in an acute leukemia mouse model. PLoS One. 2012;7(6):e37971.
17. Bindels LB, Neyrinck AM, Claus SP, et al. Synbiotic approach restores intestinal homeostasis and prolongs survival in leukaemic mice with cachexia. ISME J. 2016;10(6):1456-1470.
18. Argiles JM, Lopez-Soriano FJ, Stemmler B, Busquets S. Cancer-associated cachexia - understanding the tumour macroenvironment and microenvironment to improve management. Nat Rev Clin Oncol. 2023;20(4):250-264.
19. Campelj DG, Timpani CA, Rybalka E. Cachectic muscle wasting in acute myeloid leukaemia: a sleeping giant with dire clinical consequences. J Cachexia Sarcopenia Muscle. 2022;13(1):42-54.
20. Hoebeeck LI, Rietzschel ER, Langlois M, et al. The relationship between diet and subclinical atherosclerosis: results from the Asklepios Study. Eur J Clin Nutr. 2011;65(5):606-613.
21. Falony G, Joossens M, Vieira-Silva S, et al. Population-level analysis of gut microbiome variation. Science. 2016;352(6285):560-564.
22. Heini AD, Hugo R, Berger MD, Novak U, Bacher U, Pabst T. Simple acute phase protein score to predict long-term survival in patients with acute myeloid leukemia. Hematol Oncol. 2020;38(1):74-81.
23. Chen WL, Wang JH, Zhao AH, et al. A distinct glucose metabolism signature of acute myeloid leukemia with prognostic value. Blood. 2014;124(10):1645-1654.

24. Binder S, Luciano M, Horejs-Hoeck J. The cytokine network in acute myeloid leukemia (AML): A focus on pro- and anti-inflammatory mediators. Cytokine Growth Factor Rev. 2018;43(8-15.
25. Johnen H, Lin S, Kuffner T, et al. Tumor-induced anorexia and weight loss are mediated by the TGF-β superfamily cytokine MIC-1. Nature Medicine. 2007;13(11):1333-1340.
26. Oost LJ, Kustermann M, Armani A, Blaauw B, Romanello V. Fibroblast growth factor 21 controls mitophagy and muscle mass. J Cachexia Sarcopenia Muscle. 2019;10(3):630-642.
27. Fragkos KC, Forbes A. Citrulline as a marker of intestinal function and absorption in clinical settings: A systematic review and meta-analysis. United European Gastroenterol J. 2018;6(2):181-191.
28. Dodd D, Spitzer MH, Van Treuren W, et al. A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. Nature. 2017;551(7682):648-652.
29. Gutierrez-Diaz I, Fernandez-Navarro T, Salazar N, et al. Could Fecal Phenylacetic and Phenylpropionic Acids Be Used as Indicators of Health Status? J Agric Food Chem. 2018;66(40):10438-10446.
30. Yang QJ, Zhao JR, Hao J, et al. Serum and urine metabolomics study reveals a distinct diagnostic model for cancer cachexia. J Cachexia Sarcopenia Muscle. 2018;9(1):71-85.
31. Trombetti S, Cesaro E, Catapano R, et al. Oxidative Stress and ROS-Mediated Signaling in Leukemia: Novel Promising Perspectives to Eradicate Chemoresistant Cells in Myeloid Leukemia. Int J Mol Sci. 2021;22(5):
32. Uchiyama J, Akiyama M, Hase K, Kumagai Y, Kim YG. Gut microbiota reinforce host antioxidant capacity via the generation of reactive sulfur species. Cell Rep. 2022;38(10):110479.
33. Rashidi A, Ebadi M, Rehman TU, et al. Altered microbiota-host metabolic cross talk preceding neutropenic fever in patients with acute leukemia. Blood Adv. 2021;5(20):3937-3950.
34. Hoyles L, Fernandez-Real JM, Federici M, et al. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. Nat Med. 2018;24(7):1070-1080.
35. Vangipurapu J, Fernandes Silva L, Kuulasmaa T, Smith U, Laakso M. Microbiota-Related Metabolites and the Risk of Type 2 Diabetes. Diabetes Care. 2020;43(6):1319-1325.
36. Brial F, Chilloux J, Nielsen T, et al. Human and preclinical studies of the host-gut microbiome co-metabolite hippurate as a marker and mediator of metabolic health. Gut. 2021;70(11):2105-2114.

37. Dewulf EM, Cani PD, Claus SP, et al. Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women. Gut. 2013;62(8):1112-1121.
38. Cheng L, Hu Y, Sun J, Zhou M, Jiang Q. DincRNA: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. Bioinformatics. 2018;34(11):1953-1956.
39. Kleiman SC, Watson HJ, Bulik-Sullivan EC, et al. The Intestinal Microbiota in Acute Anorexia Nervosa and During Renourishment: Relationship to Depression, Anxiety, and Eating Disorder Psychopathology. Psychosom Med. 2015;77(9):969-981.
40. Ye H, Adane B, Khan N, et al. Subversion of Systemic Glucose Metabolism as a Mechanism to Support the Growth of Leukemia Cells. Cancer Cell. 2018;34(4):659-673 e656.
41. Szefel J, Kruszewski WJ, Ciesielski M, et al. L-carnitine and cancer cachexia. I. L-carnitine distribution and metabolic disorders in cancer cachexia. Oncol Rep. 2012;28(1):319-323.
42. Ghosh TS, Rampelli S, Jeffery IB, et al. Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: the NU-AGE 1-year dietary intervention across five European countries. Gut. 2020;69(7):1218-1228.
43. Hueso T, Ekpe K, Mayeur C, et al. Impact and consequences of intensive chemotherapy on intestinal barrier and microbiota in acute myeloid leukemia: the role of mucosal strengthening. Gut Microbes. 2020;12(1):1800897.
44. Hueso T, Gauthier J, Joncquel Chevalier-Curt M, et al. Association Between Low Plasma Level of Citrulline Before Allogeneic Hematopoietic Cell Transplantation and Severe Gastrointestinal Graft vs Host Disease. Clin Gastroenterol Hepatol. 2018;16(6):908-917 e902.
45. Ni Y, Lohinai Z, Heshiki Y, et al. Distinct composition and metabolic functions of human gut microbiota are associated with cachexia in lung cancer patients. ISME J. 2021;15(11):3207-3220.
46. Dai Z, Coker OO, Nakatsu G, et al. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. Microbiome. 2018;6(1):70.
47. Atarashi K, Suda W, Luo C, et al. Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. Science. 2017;358(6361):359-365.

48. Zeng MY, Inohara N, Nunez G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. Mucosal Immunol. 2017;10(1):18-26.
49. Bindels LB, Neyrinck AM, Loumaye A, et al. Increased gut permeability in cancer cachexia: mechanisms and clinical relevance. Oncotarget. 2018;9(26):
50. Wang R, Yang X, Liu J, et al. Gut microbiota regulates acute myeloid leukaemia via alteration of intestinal barrier function mediated by butyrate. Nat Commun. 2022;13(1):2522.

**Table 1. Study participants display similar characteristics at baseline**

| Characteristics | Total (n = 60) | CT (n = 30) | AML (n = 30) |
|---|---|---|---|
| **Age[a], *y*** | 59.0 [51.0-67.8] | 60.5 [50.0-67.3] | 59.0 [52.8-68.8] |
| **Sex, *%*** | | | |
| • **Female** | 46.7 | 46.7 | 46.7 |
| • **Male** | 53.3 | 53.3 | 53.3 |
| **BMI, *kg/m²*** | 25.0 (3.2) | 25.0 (3.2) | 25.1 (3.2) |
| **Lean mass, *%*** | 70.7 (8.4) | 71.6 (8.1) | 69.7 (8.7) |
| **Fat mass, *%*** | 25.6 (8.8) | 24.6 (8.5) | 26.6 (9.1) |
| **Smoker, *%*** | 15 | 13.3 | 16.7 |
| **Alcohol per day[a], *g*** | 8.6 [0.0-20.0] | 8.2 [2.5-18.6] | 9.7 [0.0-23.1] |
| **Overall dietary score** | 57.3 (11.2) | 59.0 (10.4) | 55.4 (12.0) |
| **Dietary quality score**** | 62.0 (18.6) | 68.8 (14.9) | 54.5 (19.6) |
| **Dietary diversity score** | 43.1 (18.2) | 40.7 (16.6) | 45.7 (19.8) |
| **Dietary equilibrium score** | 66.9 (11.0) | 67.6 (10.6) | 66.2 (11.5) |
| **(A) Adequacy score** | 71.4 (12.2) | 70.6 (10.7) | 72.4 (13.8) |
| **(B) Moderation score** | 92.6 (8.5) | 94.1 (5.9) | 91.0 (10.6) |

Variables that are normally distributed are expressed as mean (standard deviation) and are tested using a Student t-test. [a]Variables that are non-normally distributed are expressed as median (interquartile range) and are tested by a Mann-Whitney U-test. Frequency distributions are tested using a Fisher's exact test. Daily alcohol intake is missing for one AML patient (n = 29). Dietary scores are missing for 3 AML patients (n = 27). **p-value < 0.01. CT : control subjects; AML : acute myeloid leukaemia patients; BMI : body mass index.

**Figure legends**

**Figure 1.** AML (acute myeloid leukaemia) patients display alterations in the gut microbiota composition and function compared to CT (control) subjects (results of the metagenomics sequencing). A) Principal component analysis (PCA) at the genus level (metagenomics results). PERMANOVA (Permutational multivariate analysis of variance): $R^2$ = 3.1%, p-value < 0.01. B) Contribution of disease, BMI (body mass index), muscle strength, hemoglobin, sex, and age to the variance in the PCA at the genus level (PERMANOVA results). **p-value < 0.01; $^{\$}$p-value = 0.055. C) Significantly changed taxa at the lowest taxonomical level from metagenomics results. Mann-Whitney U-tests with an FDR (false discovery rate) correction were applied. Data expressed as median (interquantile range). *q-value < 0.1. D) Oral species and obligate anaerobe genera. **p-value < 0.01; *p-value < 0.05. E) PCA on bacterial EC (Enzyme Commission) enzyme functions. PERMANOVA: $R^2$ = 2.1% *ns*. F) Significantly changed bacterial EC enzyme functions in CT subjects and AML patients. *n* = 30. AML patients in orange *vs.* CT subjects in grey.

**Figure 2.** AML (acute myeloid leukaemia) patients display anorexia, muscle weakness and glycaemic disorders compared to CT (control) subjects. A) CRP (C-reactive protein), albumin and mGPS (modified Glasgow prognostic score). B) Appetite (SNAQ score, Simplified Nutritional Appetite Questionnaire) and muscle strength. C) Glycaemia (fasted), insulin and HOMA-IR2 (second homeostatic model assessment for insulin resistance). A, B : *n* = 30. C : fasted glycaemia, *n* = 20; insulin and HOMA-IR2 : *n* = 19. AML patients in orange *vs.* CT subjects in grey. Variables that are normally distributed are expressed as mean (standard deviation) and are tested using a Student t-test or a Welch's t test. Variables that are non-normally distributed are expressed as median (interquartile range) and are tested by a Mann-Whitney U-test. mGPS (modified Glasgow prognostic score) differences is tested using a $\chi^2$ test. *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001

**Figure 3.** AML (acute myeloid leukaemia) patients display inflammation and signs of gut dysfunction compared to CT (control) subjects. A) Inflammatory markers in CT subjects and AMLpatients. IL6 : interleukin-6; IL8 : interleukin-8; IL10 : interleukin-10; MCP1 : monocyte chemoattractant protein 1; TNFα : tumor necrosis factor alpha-1; TGFβ1 : transforming growth

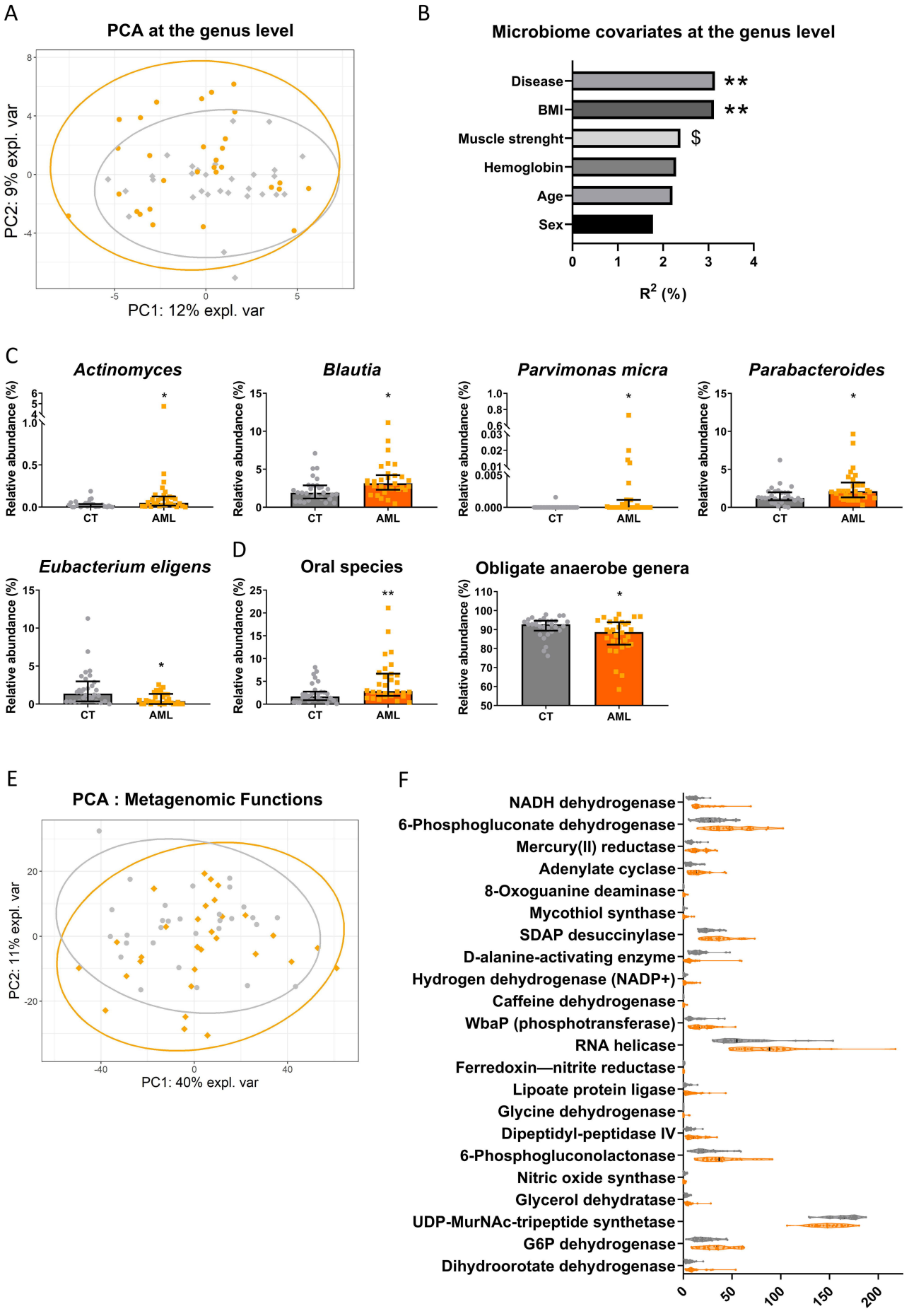
factor beta-1. B) Metabolic markers in CT subjects and AML patients. GDF15 : growth differentiation factor 15; FGF21 : fibroblast growth factor 21. C) Gut function markers in CT subjects and AML patients. LBP : lipopolysaccharide-binding protein. Variables that are normally distributed are expressed as mean (standard deviation) and are tested using a Student t-test or a Welch's t test. Variables that are non-normally distributed are expressed as median (interquartile range) and are tested by a Mann-Whitney U-test. AML patients in orange *vs.* CT subjects in grey. * p-value < 0.05; **p-value < 0.01; ***p-value < 0.001; ****p-value < 0.0001

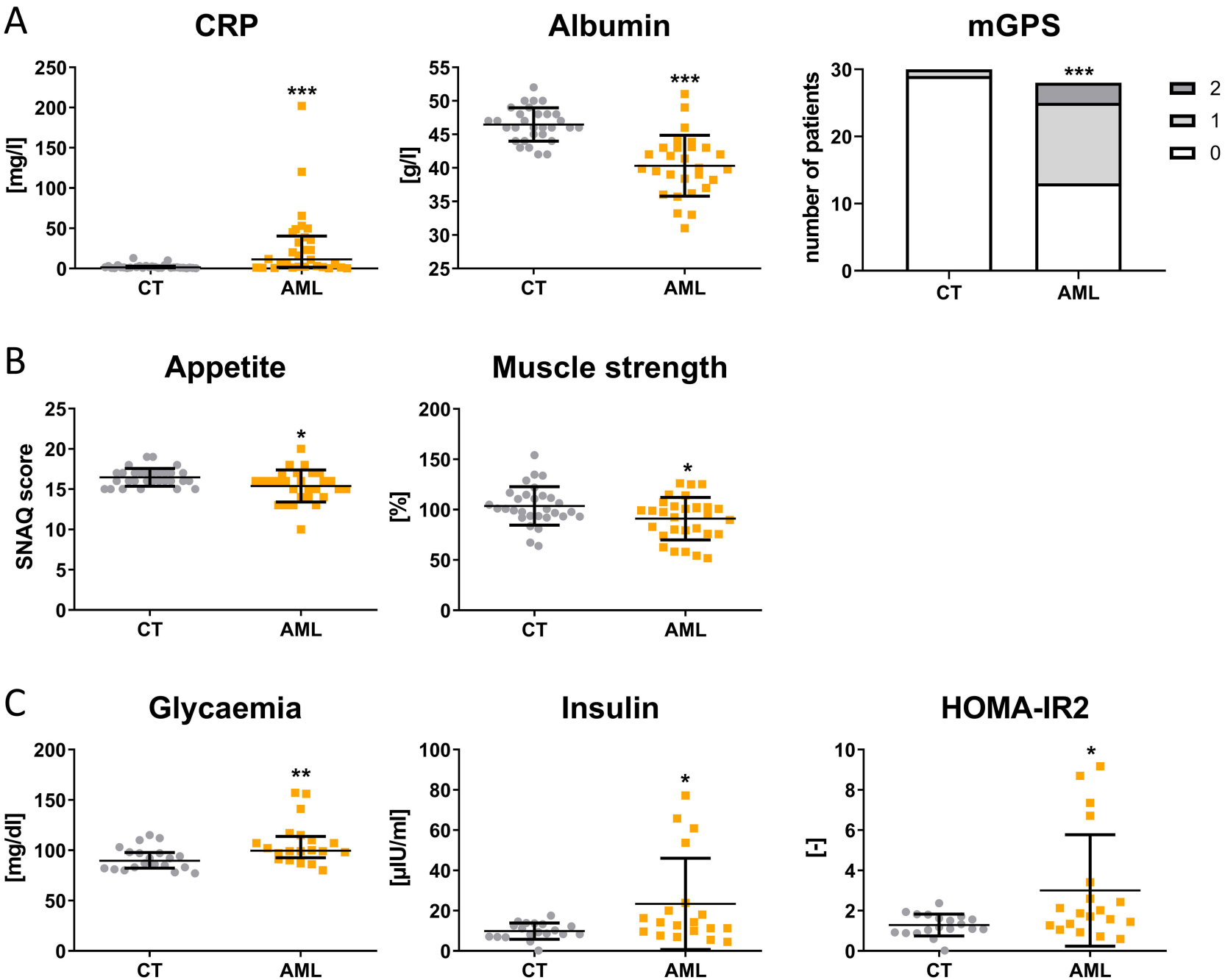
**Figure 4.** Univariate analyses pinpoint differences in the relative concentration of identified metabolites in the three analysed compartments. Bubbleplot. Bubble size depicts concentration fold change based on the group median. Coloured bubbles correspond to affected metabolites. Light and dark orange are used for metabolites increased in the AML (acute myeloid leukaemia) group (respectively at p-value < 0.05 and FDR (false discovery rate) -corrected q-value < 0.1). Light and dark grey are used for metabolites decreased in the AML group (respectively at p-value < 0.05 and FDR-corrected q-value < 0.1). Uncoloured bubbles represent non-affected metabolites. $n$ = 30 per group.

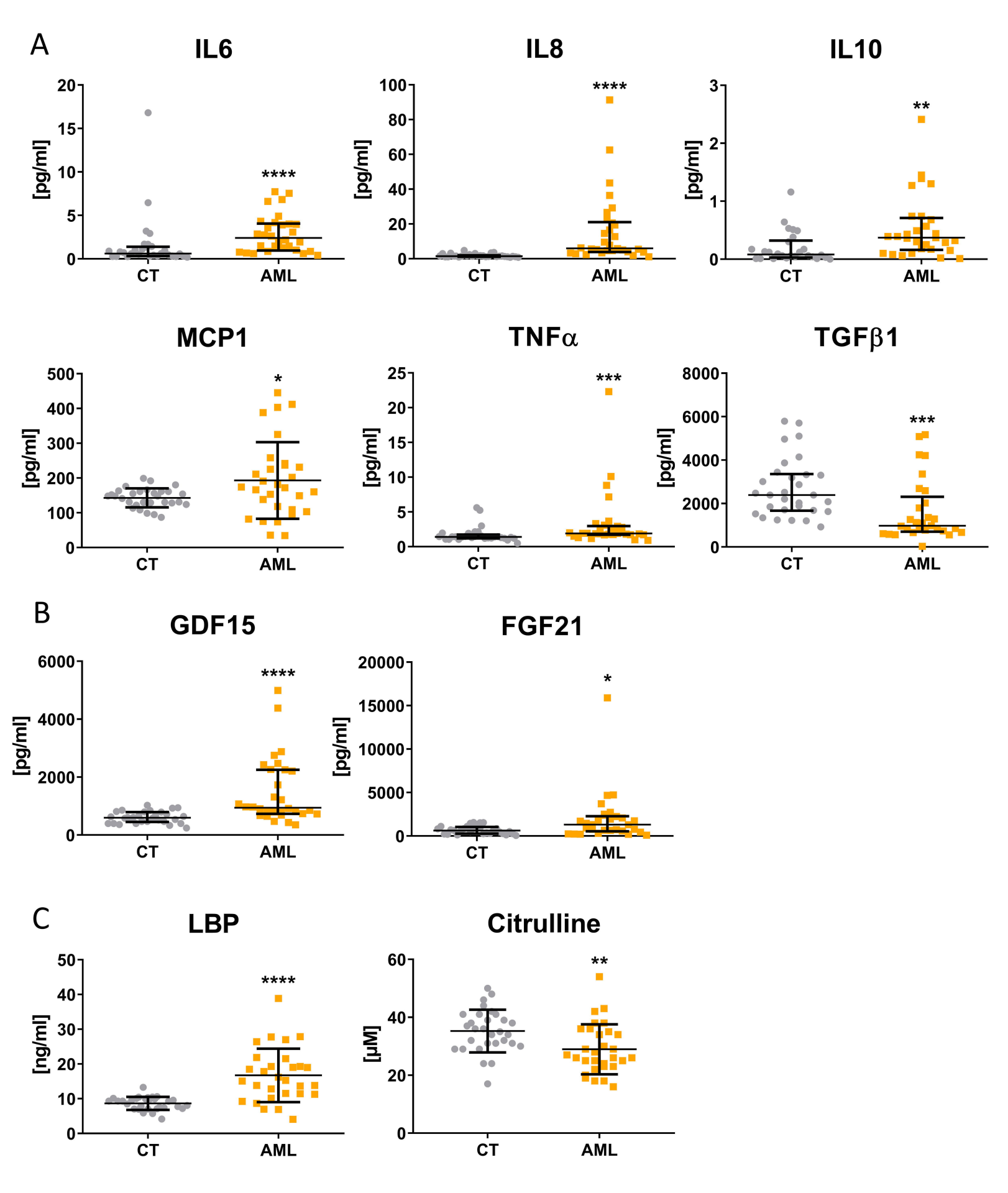
**Figure 5.** The top altered bacteria correlate with several blood and faecal metabolites. Spearman correlations. Metabolites with at least one correlation with an altered taxon are present. Microbial taxa are ordered by fold change. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR (false discovery rate) -corrected q-value < 0.1.

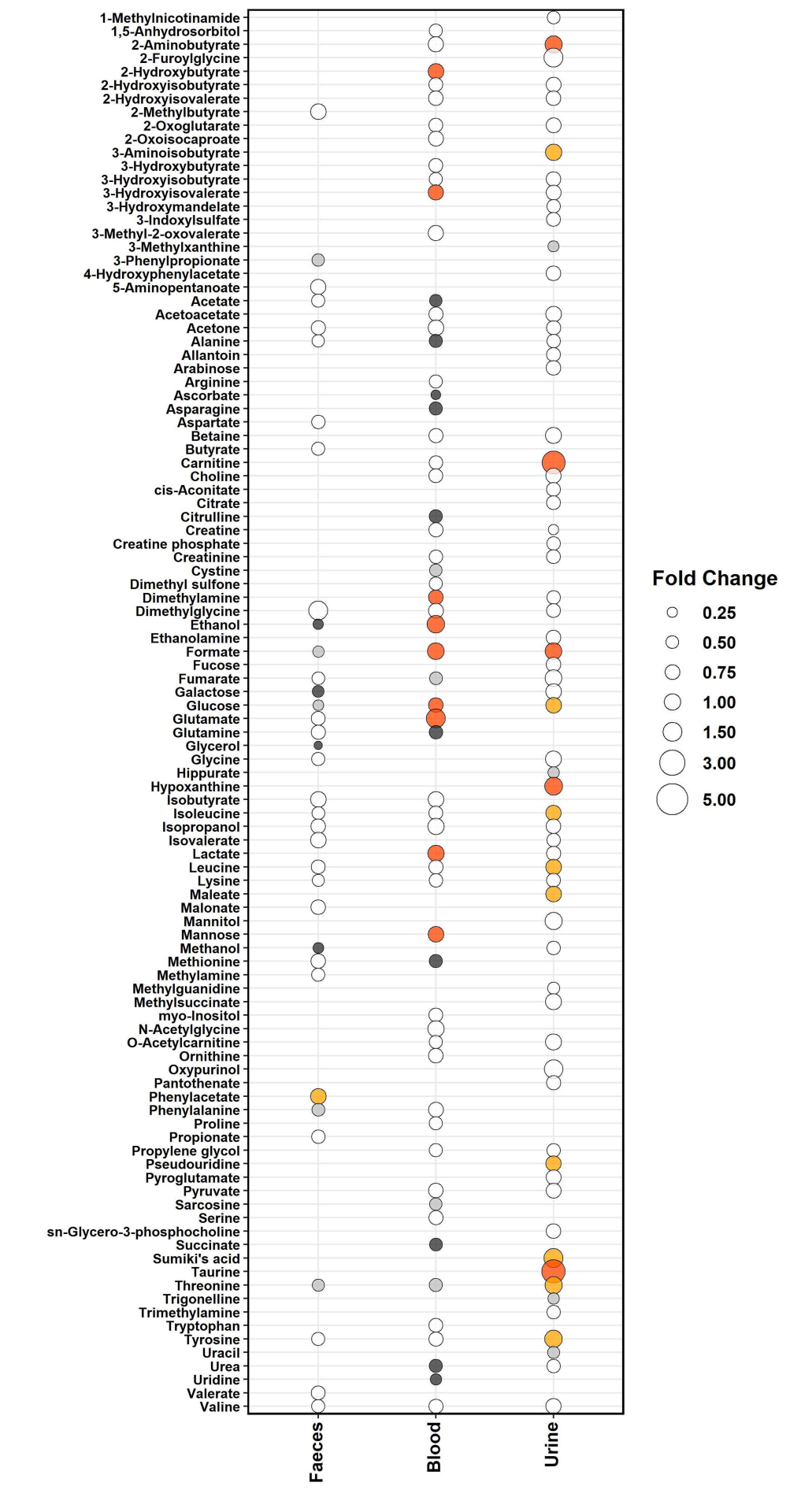
**Figure 6.** Top altered bacteria correlate with clinical, dietary, inflammatory, and metabolic parameters in CT (control) subjects and AML (acute myeloid leukaemia) patients. Spearman correlations. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR (false discovery rate) -corrected q-value < 0.1. Parameters with at least one correlation with an altered taxon are present. Microbial taxa are ordered by fold change. BMI: body mass index; WBCC: white blood cell count; appetite (SNAQ score); CRP: C-reactive protein; mGPS: modified Glasgow prognostic score; HOMA-IR2: second homeostatic model assessment for insulin resistance; IL6: interleukin-6; IL8: interleukin-8; IL10: interleukin-10; MCP1: monocyte chemoattractant protein 1; TNFα: tumor necrosis factor alpha-1; TGFβ1: transforming growth factor beta-1; GDF15: growth differentiation factor 15; FGF21: fibroblast growth factor 21; LBP: lipopolysaccharide binding protein.

**Figure 7.** Graphical abstract. Altered bacteria and functions in AML (acute myeloid leukaemia) patients associate with cachectic hallmarks and altered host redox status. Full two-way arrows display significant correlations, dashed two-way arrows display associations and simple way arrows display contributions based on literature. Created with BioRender.com.


A
PCA at the genus level
PC2: 9% expl. var
PC1: 12% expl. var
B
Microbiome covariates at the genus level
Disease
BMI
Muscle strenght
Hemoglobin
Age
Sex
**
**
$
R2 (%)
C
Actinomyces
Blautia
Parvimonas micra
Parabacteroides
Relative abundance (%)
CT
AML
D
Eubacterium eligens
Oral species
Obligate anaerobe genera
E
PCA : Metagenomic Functions
PC2: 11% expl. var
PC1: 40% expl. var
F
NADH dehydrogenase
6-Phosphogluconate dehydrogenase
Mercury(II) reductase
Adenylate cyclase
8-Oxoguanine deaminase
Mycothiol synthase
SDAP desuccinylase
D-alanine-activating enzyme
Hydrogen dehydrogenase (NADP+)
Caffeine dehydrogenase
WbaP (phosphotransferase)
RNA helicase
Ferredoxin—nitrite reductase
Lipoate protein ligase
Glycine dehydrogenase
Dipeptidyl-peptidase IV
6-Phosphogluconolactonase
Nitric oxide synthase
Glycerol dehydratase
UDP-MurNAc-tripeptide synthetase
G6P dehydrogenase
Dihydroorotate dehydrogenase

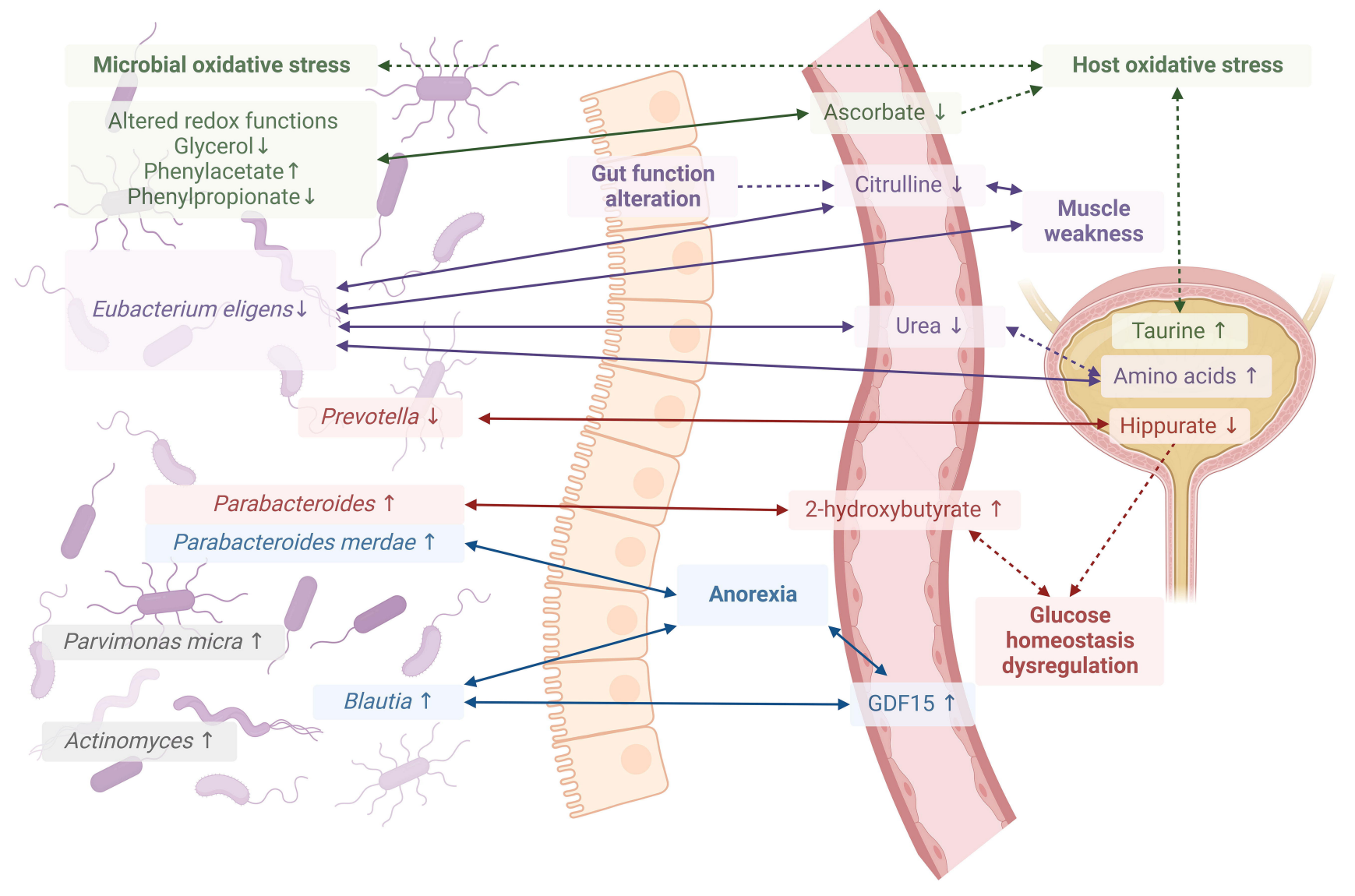

A
CRP
[mg/l]
250
200
150
100
50
0
***
CT
AML
Albumin
[g/l]
55
50
45
40
35
30
25
***
CT
AML
mGPS
number of patients
30
20
10
0
***
2
1
0
CT
AML
B
Appetite
SNAQ score
25
20
15
10
5
0
*
CT
AML
Muscle strength
[%]
200
150
100
50
0
*
CT
AML
C
Glycaemia
[mg/dl]
200
150
100
50
0
**
CT
AML
Insulin
[µIU/ml]
100
80
60
40
20
0
*
CT
AML
HOMA-IR2
[-]
10
8
6
4
2
0
*
CT
AML


A
IL6
[pg/ml]
20
15
10
5
0
****
CT
AML
IL8
[pg/ml]
100
80
60
40
20
0
****
CT
AML
IL10
[pg/ml]
3
2
1
0
**
CT
AML
MCP1
[pg/ml]
500
400
300
200
100
0
*
CT
AML
TNFα
[pg/ml]
25
20
15
10
5
0
***
CT
AML
TGFβ1
[pg/ml]
8000
6000
4000
2000
0
***
CT
AML
B
GDF15
[pg/ml]
6000
4000
2000
0
****
CT
AML
FGF21
[pg/ml]
20000
15000
10000
5000
0
*
CT
AML
C
LBP
[ng/ml]
50
40
30
20
10
0
****
CT
AML
Citrulline
[µM]
60
40
20
0
**
CT
AML


1-Methylnicotinamide
1,5-Anhydrosorbitol
2-Aminobutyrate
2-Furoylglycine
2-Hydroxybutyrate
2-Hydroxyisobutyrate
2-Hydroxyisovalerate
2-Methylbutyrate
2-Oxoglutarate
2-Oxoisocaproate
3-Aminoisobutyrate
3-Hydroxybutyrate
3-Hydroxyisobutyrate
3-Hydroxyisovalerate
3-Hydroxymandelate
3-Indoxylsulfate
3-Methyl-2-oxovalerate
3-Methylxanthine
3-Phenylpropionate
4-Hydroxyphenylacetate
5-Aminopentanoate
Acetate
Acetoacetate
Acetone
Alanine
Allantoin
Arabinose
Arginine
Ascorbate
Asparagine
Aspartate
Betaine
Butyrate
Carnitine
Choline
cis-Aconitate
Citrate
Citrulline
Creatine
Creatine phosphate
Creatinine
Cystine
Dimethyl sulfone
Dimethylamine
Dimethylglycine
Ethanol
Ethanolamine
Formate
Fucose
Fumarate
Galactose
Glucose
Glutamate
Glutamine
Glycerol
Glycine
Hippurate
Hypoxanthine
Isobutyrate
Isoleucine
Isopropanol
Isovalerate
Lactate
Leucine
Lysine
Maleate
Malonate
Mannitol
Mannose
Methanol
Methionine
Methylamine
Methylguanidine
Methylsuccinate
myo-Inositol
N-Acetylglycine
O-Acetylcarnitine
Ornithine
Oxypurinol
Pantothenate
Phenylacetate
Phenylalanine
Proline
Propionate
Propylene glycol
Pseudouridine
Pyroglutamate
Pyruvate
Sarcosine
Serine
sn-Glycero-3-phosphocholine
Succinate
Sumiki's acid
Taurine
Threonine
Trigonelline
Trimethylamine
Tryptophan
Tyrosine
Uracil
Urea
Uridine
Valerate
Valine
Faeces
Blood
Urine
Fold Change
0.25
0.50
0.75
1.00
1.50
3.00
5.00

-0.6
-0.2
0
0.2
0.4
0.6
color
Blood metabolites
Faecal metabolites
Acetoacetate
3-hydroxybutyrate
Acetone
2-oxoisocaproate
2-aminobutyrate
2-hydroxyisovalerate
Isobutyrate
3-hydroxyisovalerate
Phenylalanine
Mannose
Dimethylamine
Glucose
Isopropanol
N-acetylglycine
Creatine
Pyruvate
Glutamate
Formate
2-hydroxybutyrate
Ethanol
Urea
Citrulline
myo-Inositol
Creatinine
Propylene glycol
Tyrosine
3-hydroxyisobutyrate
2-hydroxyisobutyrate
Cystine
Lysine
Sarcosine
Alanine
2-oxoglutarate
Carnitine
Dimethyl sulfone
Acetate
Succinate
Fumarate
Uridine
Ascorbate
Isovalerate
2-methylbutyrate
Isobutyrate
Phenylacetate
Isopropanol
Methionine
Leucine
Acetone
Glutamate
Lysine
Ethanol
Galactose
Phenylalanine
Glucose
Acetate
Butyrate
Propionate
Valerate
Threonine
Glycine
Formate
Fumarate
Glycerol
Methanol
Dimethylglycine
5-aminopentanoate
Malonate
s__Prevotella_copri
f__Sutterellaceae
o__Burkholderiales
g__Prevotella
s__Eubacterium_eligens
s__Ruminococcus_bicirculans
s__Eubacterium_hallii
g__Coprococcus
s__Bilophila_wadsworthia
g__Eubacterium
g__Blautia
g__Parabacteroides
s__Parabacteroides_merdae
g__Erysipelatoclostridium
s__.Collinsella_massiliensis
s__Streptococcus_oralis
g__Lachnoclostridium
g__Actinomyces
s__Clostridium_spiroforme
g__Enorma
s__Parvimonas_micra
Decreased in AML
Increased in AML

Clinical parameters
Dietary parameters
Inflammatory and metabolic parameters
Age
Sex
BMI
Lean mass
Fat mass
Hemoglobin
WBCC
Appetite
Muscle strength
Alcohol
Fructans
Overall dietary score
Dietary quality
Dietary adequacy
Dietary moderation
CRP
Albumin
mGPS
Insulin
HOMA-IR2
IL6
IL8
IL10
MCP1
TNFalpha
TGFbeta1
GDF15
FGF21
LBP
Citrulline
s__Prevotella_copri
f__Sutterellaceae
o__Burkholderiales
g__Prevotella
s__Eubacterium_eligens
s__Ruminococcus_bicirculans
s__Eubacterium_hallii
g__Coprococcus
s__Bilophila_wadsworthia
g__Eubacterium
g__Blautia
g__Parabacteroides
s__Parabacteroides_merdae
g__Erysipelatoclostridium
s__.Collinsella._massiliensis
s__Streptococcus_oralis
g__Lachnoclostridium
g__Actinomyces
s__Clostridium_spiroforme
g__Enorma
s__Parvimonas_micra
-0.4 -0.2 0 0.2 0.4
Spearman's *rho*
Decreased in AML
Increased in AML


Microbial oxidative stress
Host oxidative stress
Altered redox functions
Glycerol ↓
Phenylacetate ↑
Phenylpropionate ↓
Ascorbate ↓
Gut function alteration
Citrulline ↓
Muscle weakness
Eubacterium eligens ↓
Urea ↓
Taurine ↑
Amino acids ↑
Hippurate ↓
Prevotella ↓
Parabacteroides ↑
2-hydroxybutyrate ↑
Parabacteroides merdae ↑
Anorexia
Glucose homeostasis dysregulation
Parvimonas micra ↑
Blautia ↑
GDF15 ↑
Actinomyces ↑

# Gut microbiome alterations at acute myeloid leukemia diagnosis are associated with muscle weakness and anorexia

Sarah A. Pötgens[1], Violaine Havelange[2,3], Sophie Lecop[1], Fuyong Li[4,5], Audrey M. Neyrinck[1], Florence Bindels[6], Nathalie Neveux[7], Jean-Baptiste Demoulin[3], Ine Moors[8], Tessa Kerre[8], Johan Maertens[9,10], Jens Walter[11], Hélène Schoemans[9,12], Nathalie M. Delzenne[1], Laure B. Bindels[1,13]

[1]Metabolism and Nutrition Research Group, Louvain Drug Research Institute, UCLouvain, Université catholique de Louvain, Brussels, Belgium.

[2]Department of Hematology, Cliniques Universitaires Saint-Luc, UCLouvain, Université catholique de Louvain, Brussels, Belgium.

[3]Experimental Medicine Unit, De Duve Institute, UCLouvain, Université catholique de Louvain, Brussels, Belgium.

[4]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Kowloon, Hong Kong SAR, China.

[5]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada.

[6]Maison Médicale de Grez-Doiceau, Grez-Doiceau, Belgium.

[7]Clinical Chemistry Department, Cochin Hospital, Paris Centre University Hospitals, Paris, France

[8]Department of Hematology, Ghent University Hospital, Ghent University, Ghent, Belgium.

[9]Department of Hematology, University Hospitals Leuven, Leuven, Belgium.

[10]Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium

[11]Department of Medicine, School of Microbiology, APC Microbiome Ireland, University College Cork, Cork, Ireland.

[12]Department of Public Health and Primary Care, ACCENT VV, KU Leuven, Leuven, Belgium

[13]Welbio Department, WEL Research Institute, Wavre, Belgium

## Supplementary Materials and Methods

### Study objectives

This cohort study aims to investigate the composition and activity of the gut microbiota of patients newly diagnosed for acute myeloid leukaemia (AML), in relationship with their food habits and cachectic hallmarks. The recruitment for this study took place with the help of clinicians, nurses and data managers at the Saint-Luc clinics, University Hospital Leuven (Campus Gasthuisberg) and University Hospital Gent.

The primary objective was to assess the composition and activity of the gut microbiota in patients with acute myeloid leukaemia (AML) compared to matched control subjects.

Secondary objectives were the following ones: (i) to investigate correlations between the gut microbiota, cachectic hallmarks and gut microbiota-related markers in the blood (gut permeability markers, microbial compounds, microbial metabolites); (ii) to characterize the changes in the gut microbial ecosystem that are induced by chemotherapy and associated with colitis; (iii) to assess whether the composition of the gut microbiota can predict the severity of chemotherapy-related colitis. Only the first secondary outcome is presented in the current manuscript.

The study was registered at ClinicalTrials.gov (NCT03881826).

### Study design

Thirty patients newly diagnosed with AML were recruited between December 2015 and December 2019 from three Belgian University hospitals (Saint-Luc Brussels (n = 13), UZ Leuven (n = 15), and UZ Gent (n = 2)). This is an academic multi-centric prospective study. Patients were included before any chemotherapy. Biological samples (urine, faeces, blood) were collected, alongside information on nutritional habits, appetite and medical records. Muscle strength and body composition were also measured. Only patients receiving a standard chemotherapy were followed after the start of the chemotherapy. For these patients, biological samples were collected and body composition, muscle strength and appetite were evaluated at 2 different time points, namely at the end of the chemotherapy and at discharge. Control (CT) subjects from the general population were recruited between December 2017 and January 2020 based on the same inclusion/exclusion criteria, except for the AML diagnosis. They were matched (1:1) for several factors known to impact GM, such as age[1], sex[2-4], BMI[5], and smoking status[6]. Whole-group analyses were applied

on these matched cohorts as previously advised[7]. When we initiated the project in 2015, sample size could not be calculated as the effect size and the inter-individual variability were unknown. This study was considered as exploratory and expected to provide information concerning the effect size[8]. The number of patients was therefore chosen based on similar studies[9-13]. Retrospectively, we estimated the power of the MicroAML study using information collected in a previous study performed in a cohort of 24 healthy volunteers[10, 14]. In the Food4Gut healthy cohort, we found an average standard deviation of 12% for the Shannon index of alpha-diversity[10], a measurement of the microbial diversity. Using PASS 14.0.7, we found a power of 89% to detect with 30 subjects/group at a threshold p-value (alpha) of 0.05 a minimal 10% change in this alpha-diversity index, supposing a similar standard deviation of 12%. This calculation indicates that the MicroAML study is adequately powered to detect such changes in the gut microbiota of leukemic patients *vs* healthy volunteers.

*Inclusion Criteria for AML patients*

Patients with

- A diagnosis of AML and related precursor neoplasms according to WHO 2008 classification (excluding acute promyelocytic leukaemia) including secondary AML (after an antecedent haematological disease (e.g. MDS) and therapy-related AML) OR acute leukaemia's of ambiguous lineage according to WHO 2008 OR a diagnosis of refractory anaemia with excess of blasts (MDS REAB) 2 and IPSS (International Prognostic Scoring System)-R score > 2.
- World Health Organization performance status 0, 1 or 2
- Sampled bone marrow and/ blood cells at diagnosis with molecular analysis.
- Written informed consent
- Good command of the French or Dutch language

*Inclusion Criteria for CT subjects*

- For each enrolled patient, a healthy control was recruited and matched for age, sex, BMI and smoking habits (except one).
- Written informed consent
- Good command of the French or Dutch language

*Exclusion Criteria for AML patients and CT subjects*

- Age < 18 years
- Age > 75 years
- Pregnancy
- Antibiotics consumption during the last 30 days before inclusion
- Recent chemotherapy (< 3 months), with exclusion of hydroxyurea
- BMI >30
- Any history of chronic intestinal affections (Crohn disease, inflammatory bowel disease, gluten intolerance)
- Gastric bypass
- Current treatment with antidiabetic or hypoglycaemic drugs

**Data collection**

All biological sampling and data collection were performed at the time of diagnosis before the beginning of the chemotherapy treatment and the administration of any antibiotics.

*Biological sampling*

Faeces and urine were collected and were immediately (< 15 min) frozen at -20°C for a maximum of one week and then stored at -80°C. Blood samples were kept on ice, centrifuged at 4°C within 30 min and plasma aliquots were stored at -80°C. Fasting status was reported.

*Case report form*

Case report forms were collected to document medical history, drug records including consumption of pre- and probiotics, antibiotics in the last 90 days, as well as lab assessment of haemoglobin, white blood cell count, C-reactive protein, albumin, and glycaemia levels.

*Body composition and muscle strength*

Body composition was assessed using bioimpedancemetry (Body composition analyser, Tanita BC-420MA/SMA). Muscle strength was measured using a Jamar hand dynamometer in the dominant hand (3 measures, separated by 15 s). Patients were asked for weight loss during the last six months.

*Dietary and other assessments*

The overall quality of patients' dietary habits was evaluated by a food frequency questionnaire (FFQ) validated in the Belgian population[15]. The analysis of the FFQ gives an overall dietary score and several sub-scores: dietary quality score, dietary diversity score, dietary equilibrium score (adequacy and moderation scores). Patients also filled questionnaires to evaluate their alcohol intake on a weekly basis. The short tobacco test was used to evaluate tobacco dependence and consumption[16]. Appetite was assessed using the simplified nutritional assessment questionnaire (SNAQ)[17]. A score ≤ 14 reflects a risk of weight loss in the next six months.

**Measurement of cytokines, GDF15, FGF21, LBP, insulin and citrulline**

Plasma cytokines (IL6, IL8, IL10, MCP1, TNFα, TGFβ1, GDF15), FGF21 and insulin (in fasted state) were measured using a customized U-plex kit and a Meso Scale Discovery microplate reader (Meso Scale Discovery, Rockville, MD, USA). LBP levels were assessed using an ELISA kit (HycultBiotech, PA, USA). Citrulline was measured in plasma (EDTA) using ion exchange chromatography. Combining fasted glycaemia and insulin, we calculated the HOMA-IR2 index[18] for 19 patients in each group.

**Gut microbiome analyses**

*DNA extraction and total bacteria quantification*

DNA was extracted from faecal samples following the protocol Q described by Costea et al[19]. This protocol uses the QIAamp DNA Stool Mini Kit (Qiagen, Germany) and includes a bead-beating step. Treatment with RNAse A was performed (10 mg/ml, Thermo Fisher Scientific, USA). DNA concentration was determined, and purity (A260/A280) was checked using a NanoDrop 2000 (Thermo Fisher Scientific, USA).

Absolute quantification of the total bacterial load was performed by quantitative polymerase chain reaction (qPCR) using the primers Bacteria Universal P338F (ACTCCTACGGGAGGCAGCAG) and P518r (ATTACCGCGGCTGCTGG)[20]. Real-time PCR was performed with a QuantStudio3 (Applied Biosystems, The Netherlands) using SYBR Green (GoTaq® qPCR mix, Promega, USA) for detection. All samples (0.1ng/µl) were run in duplicate in a single 96-well reaction plate. Final concentrations were as follow: cDNA 2 µl/25 µl, primers 300 nM, and SyberGreen mix 1X (MeteorTaq DNA polymerase, dNTP, RT buffer, $MgCl_2$ 4 mM, SYBR® Green I, ROX passive reference and stabilizers, as provided by the manufacturer). Thermocycling

conditions were as follow: initiation step at 95°C 2 min; cycling stage at 95°C 30 s, 60°C 30 s, 72°C 30 s, 40 cycles; melt curve stage at 95°C 1 s, 65°C 20 s, increment of 0.1°C every 1 s until reaching 95°C. Threshold was manually adjusted to reach the linear range of the log-fluorescent curves and CT values were determined using the QuantStudio Software (Version 1.4.3, Applied Biosystems, The Netherlands). Absolute quantification was achieved through the inclusion of a standard curve (performed in duplicate) on each plate generated by diluting DNA from pure culture of *L. acidophilus* NCFM (five-fold serial dilution). Cell counts were determined by plating and expressed as "colony-forming unit" (CFU) before DNA isolation.

*16S rRNA gene sequencing - data generation*

Sequencing of 16S rRNA gene is a well-established technique allowing taxonomical assessment of the gut microbiota. This method uses primers that target a specific region of the 16S rRNA gene. Indeed, this gene has the advantage to have highly variable regions flanked by highly conserved regions in all bacteria. The sequencing of these variable regions allows microbial phylogenies determination. In this study, amplicon sequencing of the microbiome was done at the University of Minnesota Genomics Centre. Briefly, the V5-V6 region of the 16S rRNA gene was PCR-enriched using the primer pair V5F_Nextera (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGRGGATTAGATACCC) and V6R_Nextera (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGACRRCCATGCANCACCT) in a 25 µl PCR reaction containing 5 µl of template DNA, 5 µl of 2X HotStar PCR master mix, 500 nM of final concentration of primers and 0.025 U/µl of HostStar Taq+ polymerase (QIAGEN). PCR-enrichment reactions were conducted as follow, an initial denaturation step at 95°C for 5 min followed by 25 cycles of denaturation (20 s at 98°C), annealing (15 s at 55°C), and elongation (1 min at 72°C), and a final elongation step (5 min at 72°C). Next, the PCR-enriched samples were diluted 1:100 in water for input into library tailing PCR. The PCR reaction was analogous to the one conducted for enrichment except with a KAPA HiFi Hot Start Polymerase concentration of 0.25 U/µl, while the cycling conditions used were as follows: initial denaturation at 95°C for 5 min followed by 10 cycles of denaturation (20 s at 98°C), annealing (15 s at 55°C), and elongation (1 min at 72°C), and a final elongation step (5 min at 72°C). The primers used for tailing are the following: F-indexing primer AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC and R-indexing

primer CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGG GCTCGG, where [i5] and [i7] refer to the index sequence codes used by Illumina. The resulting 10 µl indexing PCR reactions were normalized using a SequalPrep normalization plate according to the manufacturer's instructions (Life Technologies). 20 µl of each normalized sample was pooled into a trough, and a SpeedVac was used to concentrate the sample pool down to 100 µl. The pool was then cleaned using 1X AMPureXP beads and eluted in 25 µl of nuclease-free water. The final pool was quantitated by QUBIT (Life Technologies) and checked on a Bioanalyzer High-Sensitivity DNA Chip (Agilent Technologies) to ensure correct amplicon size. The final pool was then normalized to 2 nM, denatured with NaOH, diluted to 8 pM in Illumina's HT1 buffer, spiked with 20% PhiX, and heat denatured at 96°C for 2 min immediately prior to loading. A MiSeq 600 cycle v3 kit was used to sequence the pool. Raw sequences can be found in the SRA database (project ID: PRJNA813705).

*16S rRNA gene sequencing - bioinformatics*

Subsequent bioinformatics analyses were performed *in-house* as previously described[21]. Initial quality filtering of the reads was performed with the Illumina Software, yielding an average of 84 585 pass-filter reads per sample. Quality scores were visualized with the FastQC software (http://www.bioinformatics.babraham.ac.uk/ publications.html), and reads were trimmed to 220 bp (R1) and 200 bp (R2) with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Next, reads were merged with the merge-illumina-pairs application v1.4.2 (with P = 0.03, enforced Q30 check, perfect matching to primers which are removed by the software, and otherwise default settings including no ambiguous nucleotides allowed)[22]. The UPARSE pipeline implemented in USEARCH v11[23] was used to further process the sequences. Amplicon sequencing variants (ASVs) were identified using UNOISE3[24]. Such method infers the biological sequences in the sample prior to the introduction of amplification and sequencing errors, and distinguishes sequence variants differing by as little as one nucleotide[25]. The analysis allowed the identification of 3968 ASVs. ASVs were identified using the RDP database. Taxonomic prediction was performed using the *nbc_tax* function[26], an implementation of the RDP Naive Bayesian Classifier algorithm[27]. Alpha diversity indexes were calculated using QIIME[28] on the rarefied ASV table. Rarefaction was performed using Mothur 1.32.1[29] by randomly selecting 40 612 sequences for all samples, except two (103: 24 133 sequences and 507: 16 814 sequences).

*16S rRNA gene sequencing - biostatistics*

Unrarefied data were filtered to select for a minimum abundance of 0.01% and a minimal prevalence of 25% in one group. Principal component analysis (PCA) was performed on CLR-transformed data[30] using the *pca* function in the *mixOmics* R package[31]. The CLR transformation consists in a centered log ratio transformation and allows transforming compositional data into an Euclidian space. A pseudo-count equal to half the minimal value found in the dataset was applied prior the CLR transformation[32]. Significantly impacted phyla, families and genera were identified using a Mann-Whitney U-test in R since normality was not inspected for every phylum/family/genus. The p-value was adjusted to control for the false discovery rate (FDR) for multiple testing according to the Benjamini and Hochberg (BH) procedure[33]. A q-value < 0.1 was considered significant.

*Metagenomics sequencing - data generation*

Untargeted metagenomics sequencing was performed at the Centre d'expertise et de services Génome Québec. Genomic DNA was quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies). Libraries were generated from 50 ng of genomic DNA using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) as per the manufacturer's recommendations. Adapters and PCR primers were purchased from IDT. Size selection of libraries contained the desired insert size has been performed using SparQ beads (Qiagen). Libraries were quantified using the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). Average size fragment was determined using a LabChip GXII (PerkinElmer) instrument.

The libraries were normalized and pooled and then denatured in 0.05 N NaOH and neutralized using HT1 buffer. The pool was loaded at 225 pM on an Illumina NovaSeq S4 lane using Xp protocol as per the manufacturer's recommendations. The run was performed for 2x150 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at 1% level. Base calling was performed with RTA v3.4.4 . Program bcl2fastq2 v2.20 was then used to demultiplex samples and generate fastq reads.

*Metagenomics sequencing – bioinformatics*

Trimmomatic (version 0.39)[34] was used to trim adapters and low-quality reads (average quality scores < 20), and only reads with the length no less than 100 bp remained for the downstream

analysis. Bowtie2 (version 2.3.5.1)[35], with *-N 1* and otherwise default options, was applied to remove reads classified as bacteriophage phiX174 (NCBI accession: NC_001422.1) and filter out human DNA reads based on the human genome reference GRCh38 . MetaPhlAn3 (version 3.0.2) and HUMAanN 3.0 was used to estimate the taxonomic composition and functional profiles of the gut microbiome, with the default settings[36]. Genes were then regrouped in 2373 Level-4 enzyme commission (EC) categories system (*humann_regroup_table --groups uniref90_level4ec*). Both genes and EC enzyme functions were normalized in cpm. The *human_barplot* function was used to explore the contribution of individual species and genera to selected functions. Raw sequences can be found in the SRA database (project ID: PRJNA813705).

*Metagenomics sequencing – biostatistics on taxonomical data*

Taxa were filtered to select for 320 taxa with a mean average abundance above 0.01% and a mean prevalence of 25% in at least one group of samples. PCA were computed from CLR-transformed data[30] followed by Permutational Multivariate Analysis of Variance (PERMANOVA) using the *adonis* function in the *vegan* R package[37]. Different variables were tested, including AML, BMI, sex, muscle strength, haemoglobin and age. The PERMANOVA allowed to evaluate the explanatory power of each factor individually. Significantly affected taxa were identified using a Mann-Whitney U-test with BH correction. A q-value < 0.1 was considered significant.

As multiple differential abundance methods help to ensure robust biological interpretation[38], we also used alternative differential abundance methods, namely a Mann-Whitney U-test with BH correction applied on CLR-transformed data and ALDEx2[39]. The R scripts used to perform differential abundance analyses are available on GitHub at the following address: https://github.com/laurebindels/MicroAML. Similar results were found using these methods. A list of bacteria of putative interest was gathered for data integration with metabolomics datasets as well as clinical and biochemical data, by selecting bacterial taxa present in the top 20 of each approach. When parent taxa were present in this list and identical/highly similar in terms of abundance values, the lowest taxonomical level was conserved. This approach allowed the selection of a list of 21 taxa referred in the manuscript as “top altered bacteria” and are presented in Table S2.

Bacterial features were estimated as previously described[40]. For each sample, the cumulative relative abundance of taxa that were associated with an obligate anaerobic metabolism or an oral habitat was determined. The level of oral bacteria was computed based on an aggregation at the

species level and the expanded Human oral microbiome database V3[41]. The level of obligate anaerobe bacteria was computed based on an aggregation at the genus level and the oxygen class indicated in the List of Prokaryotes according to their Aerotolerant or Obligate Anaerobic Metabolism (OXYTOL 1.3, Mediterranean institute of infection in Marseille).

The Random Forest algorithm was used to model the bacterial taxonomic signature of the AML status. The AUC or "area under the receiver-operator curve" measures the accuracy of trained forests. The AUC is a widely used estimator of true positive and false positive prediction rates. For this analysis, outcomes were AML or no AML and the dataset, namely the relative abundance of taxa identified in the top altered bacteria, was randomly split in a training and a testing set with a ratio of 0.666 (20 patients in the training dataset and 10 patients in the testing dataset) using the *caTools* R package[42]. Using the *randomForest*[43] and *ROCR*[44] R packages, we trained 300 forests, containing 1001 trees each, with the training dataset, and we selected the model with the highest AUC. The accuracy of this model was predicted using the testing dataset. A trained forest produces a variable importance list based on mean decrease accuracy. For this analysis, the variable importance list is a list of taxa that contributed most to the correct group assignment of every sample and is presented in Figure S5.

*Metagenomics sequencing – biostatistics on functional data*

Functions were filtered to select for 1465 functions with a mean average abundance above 1 and a mean prevalence of 25% in at least one group of samples. PCA were computed from CLR-transformed data[30] followed by PERMANOVA using the *adonis* function in the *vegan* R package[37]. Twenty-two significantly affected functions were identified using MaAsLin2[36] (LM method, LOG transformation, no normalization). Model validation was achieved for the top 5 functions by visual inspection of the plot of the residuals against the fitted values. The LM method was preferred to the CPLM method based on the distribution of the residues of the models for these top 5 significant features. A q-value < 0.1 was considered significant.

The Random Forest algorithm was used to model the bacterial functional signature of the AML status. For this analysis, outcomes were AML or no AML and the dataset, namely the relative abundance of altered bacterial functions, was randomly split in a training and a testing set with a ratio of 0.666 (20 patients in the training dataset and 10 patients in the testing dataset) using the *caTools* R package[42]. Using the *randomForest*[43] and *ROCR*[44] R packages, we trained 300 forests,

containing 1001 trees each, with the training dataset, and we selected the model with the highest AUC. The accuracy of this model was predicted using the testing dataset. A trained forest produces a variable importance list based on mean decrease accuracy. For this analysis, the variable importance list is a list of that contributed most to the correct group assignment of every sample and is presented in Figure S5.

## $^{1}$H-NMR Metabolomics analyses

*Sample preparation*

Faecal samples were prepared as follow: 200 mg of faeces were diluted into 1000 µl NMR buffer ($H_2O$–$D_2O$ (1:1), pH = 7 ($NaHPO_4$-$NaH_2PO_4$ 0.2 M), trimethylsilylpropanoic acid (TSP) 1 mM as standard) and homogenised in a TissueLyser (4 min, 25 Hz). The homogenate was centrifuged (10 min 13 000 g 4°C). The supernatant was then transferred into a 1.5 ml Eppendorf tube for a second centrifugation (3 min 13 000 g 4°C). This last supernatant was transferred into 5 mm diameter NMR tubes.

Plasma samples (Heparin tubes) were prepared as follow: AMICON ultra 0.5 ml – 10 kDa filters tubes were rinsed 5 times with 500 µl of distilled water followed by centrifugation (15 min 14 000 g 4°C). 500 µl of plasma were then filtered (30 min 14 000 g 4°C) followed by 250 µl of phosphate buffer (30 min 14 000 g 4°C) ($H_2O$–$D_2O$ (1:9), pH = 7 ($NaHPO_4$-$NaH_2PO_4$ 0.2 M). 150 µl of phosphate buffer ($D_2O$, pH = 7 ($NaHPO_4$-$NaH_2PO_4$ 0.2 M), TSP 4 mM as standard) were directly added to the filtrate. After mixing (5 s vortex) 600 µl were transferred into 5 mm NMR tubes.

Urine samples were prepared as follow: urine samples were thawed overnight at 4°C, mixed (5 s vortex) and centrifuged (5 min 3 500 g 4°C). 630 µl of the supernatant were then mixed (5 s vortex) with 70 µl of NMR buffer ($D_2O$, pH = 7.1 ($KHPO_4$-$K_2HPO_4$ 1.5 M), TSP 2.9 mM, $NaN_3$ 0.2%) and then centrifuged (5 min 3 500 g 4°C). 600 µl were transferred into 5 mm NMR tubes. pH was measured to check that all samples ranged between 6.8 and 7.1.

Note that samples were randomly allocated to four groups that were analysed on four consecutive days, so that each sample spectrum was acquired within 12 hours of its preparation. Quality controls were included in each set of samples (one per day) to ensure the reproducibility and homogeneity of the obtained data. For faeces and plasma analyses, quality controls consisted in extra aliquots of

faces/plasma of a volunteer. For urine analyses, quality controls consisted in an aliquot of a pool (four different individuals) of AML and CT urine extra samples.

*Data collection*

NMR data were acquired on a Bruker Avance 600 MHz NMR spectrometer equipped with a cryoprobe. During acquisition, sample temperature was maintained at 300 K. Spectra were collected with a 1D NOESY pulse sequence for the plasma and urine samples and with a 1D CPMG pulse sequence for faecal samples. The 1D NOESY pulse sequence covered 21 ppm. Spectra were digitized in 65K data points during a 2.6 s acquisition time. The mixing time was set to 10 ms, and the relaxation delay between scans was set to 4 s. The 1D CPMG pulse sequence covered 20 ppm. Spectra were digitized in 65K data points during a 2.7 s acquisition time. The relaxation delay between scans was set to 4 s. Spectra were acquired using 128 scans for faecal and urine samples and 256 scans for plasma samples. To confirm metabolite identification, 2D $^{1}$H-$^{1}$H NMR spectra, such as J-RES and TOCSY, as well as $^{1}$H-$^{13}$C HSQC, were acquired for selected samples.

*Data processing*

The data were processed using MestReNova (v14.2). The spectra were zero filled with a factor of two. They were submitted to apodization using a 0.3 Hz decaying exponential function and fast Fourier transformed. Automated phase correction and second-order polynomial baseline correction were applied to all samples. All spectra were aligned on TSP. Spectral quality control was performed and some spectra were re-run. After re-run, all spectra passed the quality control and were included in further analyses. Only the region from 0.12 to 10 ppm was conserved. Water signal was removed from all spectra before statistical analyses. Considering the contaminations due to the filtering step, regions 3.36-3.37, 3.55-3.58 and 3.64-3.68 ppm were excluded from the analyses for plasma samples. Probabilistic quotient normalisation was performed. Intelligent bucketing was realised using the Matlab software (v9.2)[45]. Metabolites were assigned using the Chenomx NMR Suite (v8.43), the Bruker B-BIOREFCODE database (Amix software v3.9.15), the HMDB[46] and additional 2D NMR experiments on selected representative samples. The Chenomx NMR Suite was used to perform a relative quantification of the identified metabolite concentrations. TSP was used as a chemical shift and quantification reference for all spectra. Quantitative fitting of each spectrum was carried out in batch mode, followed by manual adjustment.

*Statistical analyses*

The tables of metabolite concentrations for each compartment were analysed in R. When missing data did not exceed 40% in both groups (CT and AML), a left-censored missing data imputation method was applied using the *impute.QRILC* function implemented in the *imputeLCMD* R package[47]. This was possible for all metabolites presented in this paper except for 3-phenylpropionate and maleate. For these two metabolites, no imputation was performed and the difference between both groups was tested by a Fisher's Exact Test in R. Significantly affected metabolites were identified by a Mann–Whitney U test since normality was not tested for each metabolite. The p-value was adjusted to control for the FDR. Q-values inferior to 0.1 were kept. The bubbleplot was generated using an *in-house* script including the *tidyverse* R package[48]. PCA with scaling to unit variance and partial least square discriminant analysis (PLS-DA) were respectively performed using the functions *pca* and *plsda* of the *mixOmics* R package[31].

**Statistical analyses**

*General statistical overview*

Normality was assessed using d'Agostino and Pearson omnibus normality test. If normality was not respected in one group, the nonparametric Mann–Whitney U test was used. Fisher's exact test was used to check for variance equality between groups. Student's t-test was used when variances were not statistically different. In case of variance inequality, Welch's t-test was used. Coherently, normal variables are presented as mean with standard deviation (SD) whereas non-normal variables are presented as median with interquartile range (IQR). $P < 0.05$ was considered statistically significant. For all *-omics* data, normality was not assessed and Mann-Whitney U-tests were therefore used. When needed, a correction for FDR was applied[33]. In this case, a q-value < 0.1 was considered significant. Statistical analyses were performed using GraphPad Prism 8.0 and R.

*Integration analyses*

Correlations between the top altered bacteria and (i) faecal, blood and urine metabolites and (ii) clinical, dietary, inflammatory, and metabolic parameters were performed in R using Spearman correlations. When adjusting for potential covariates such as age, partial Spearman rank-based correlations were computed using the *pcor* function (http://www.yilab.gatech.edu/pcor). FDR correction was applied. Metabolites/parameters with at least one correlation with one of the bacteria with a p-value < 0.05 were kept for inclusion in the heatmap. P-value < 0.05 were marked with a

‘+’ and q-value < 0.1 were marked with a ‘*’. The same method was applied for the correlations between the top bacterial functions and faecal, blood and urine metabolites.

## Supplementary Results and Discussion

### AML patients with hyperinsulinemia or hyperglycaemia display specific GM alterations

When looking at the glycaemia, insulin and HOMA-IR2 results presented in Figure 2C, it appears that some AML patients present much higher level of these parameters compared to the rest of the cohort. The four patients that are very different to the rest of the AML group for insulin levels and HOMA-IR2 are the same. Among those four patients, only one of them has also a higher glycaemia than the rest of the patients. The other two patients that have higher glycaemia levels do not have insulin levels and HOMA-IR2 that differ from the rest of the AML group.

The analysis of the microbiota stratifying the individuals by insulinemia and glycaemia were performed separately. Individuals were stratified in three groups, namely CT_low, AML_low and AML_high. For the analysis stratifying by insulinemia, the four patients with higher insulin levels than the rest of the cohort were allocated to the AML_high group, the rest of the AML patients were allocated to the AML_low group and the healthy individuals were included in the CT_low group. For the analysis stratifying by glycaemia, the three patients with high glycaemia levels (defined as glycaemia above 120mg/dl) were included in the AML_high group, the rest of the AML patients were allocated to the AML_low group and the healthy individuals were included in the CT_low group.

PCA at the species level stratified by insulinemia did not show a clear clustering of the AML_high group (new Figure S7A) since all three groups were superimposed. Moreover, insulinemia class did not explain a significant part of the variance in the dataset according to the PERMANOVA analysis. Interestingly, four species were significantly different between individuals with low versus high insulin levels (p-value < 0.05, q-value ns) (Figure S7B). *Phascolarctobacterium faecium*, *Bacteroides caccae* and *Bacteroides fragilis* were more abundant in the AML_high group whereas *Eubacterium eligens* was decreased in AML_high individuals. Similarly, PCA at the species level stratified by glycaemia did not show a clear clustering of the AML_high group (Figure S8A) and the PERMANOVA analysis was not significant either. However, six species were significantly different between individuals with low versus high glycaemia (p-value < 0.05, q-value ns) (Figure S8B). *Intestinibacter bartlettii*, *Bacteroides ovatus* and *Fusicatenibacter*

*saccharivorans* were decreased in the AML_high group whereas *Clostridium sp CAG 242*, *Firmicutes bacterium CAG 94* and *Streptococcus oralis* were decreased in those same individuals compared to the rest of the cohort.

Different databases, namely Disbiome[49], gutMDisorder[50], and Pubmed, were screened to investigate whether these bacteria of interest were previously associated with diseases or syndromes characterized by high blood levels of insulin and glucose, such as type I and II diabetes, obesity and metabolic syndrome. A compilation of the results is presented in Table S5.

The increase of *Phascolarctobacterium faecium*, *Bacteroides caccae* and *Streptococcus oralis* found in the AML_high group is in line with its increased abundance in individuals with type I or II diabetes. In the same manner, the decrease of *Eubacterium eligens* and *Intestinibacter bartlettii* is in accordance with its decreased abundance in individuals with type I diabetes among other diseases. Interestingly, *Intestinibacter bartlettii* was found to be correlated with markers for insulin resistance in 53 postmenopausal women with obesity[51]. In contrast, the increase of *Bacteroides fragilis* and *Bacteroides ovatus* observed in the AML_high group is in line with the results found in some studies but not others. *Bacteroides fragilis* was found to be increased in children with type I diabetes in one study[52] and in children with obesity in another study[53]. However, *Bacteroides fragilis* was also found to be decreased in a study with children with type I diabetes[54]. The results found concerning *Bacteroides ovatus* were also contradictory. In our study, *Bacteroides ovatus* was significantly decreased in AML patients with higher glycaemia levels, such as in children with obesity[55]. However, *B. ovatus* was increased in two studies of children with type I diabetes[52, 54]. Unfortunately, no information was found concerning the abundance of *Fusicatenibacter saccharivorans*, *Clostridium sp CAG 242* and *Firmicutes bacterium CAG 94* in diseases and syndromes characterized by high insulinemia and glycaemia.

To conclude, although the results presented above should be interpreted with caution given the very limited number of patients in the AML patients with high insulinemia/glycaemia, hence the lack of significance after correction for multiple testing, the bacterial species that vary significantly appear to be consistent with the information present in the literature.

**Blood glutamine levels correlate with skeletal muscle mass and function**

The plasma metabolomic signature of AML patients revealed signs of purine nucleotide metabolism deficiency and metabolic stress (e.g., increased hypoxanthine, reduction in TCA cycle intermediates, decreased glutamine alongside increased glutamate). Whether this stress may contribute to muscle alterations was explored through an additional set of correlations (Figure S22). Among these metabolites, only glutamine significantly correlated with lean mass, lean weight, and muscle strength (Figure S22). This correlation could reflect a deleterious impact of glutamine depletion driven by AML cells on the muscle as the skeletal muscle is the main storage site and endogenous source for glutamine.

**Additional discussion on the potential of FMT in AML treatment and cachexia**

Patients with AML usually receive induction chemotherapy coupled with antibiotic treatment. Those patients experience an alteration of the gut microbiota[56, 57] that remains after the end of the treatment[58, 59].

Fecal microbiota transplant represents an actionable measure to counteract the effect of treatment on the microbiota and cachexia. As stated in the introduction, Malard and colleagues[60] investigated the safety and diversity-enhancing ability of autologous fecal microbiota transfer (FMT) in patients with AML receiving intensive chemotherapy and antibiotics. Fecal material collected at the time of diagnosis was used for fecal microbiota transfer. This transfer appeared to be safe and could restore microbial richness and diversity based on α-diversity indices. However, autologous FMT with an intact microbiota is not always possible since patients have often already received antibiotic treatment by the time of diagnosis. Third-party FMT are thus an alternative. A randomized double-blind placebo-controlled trial on allogeneic hematopoietic cell transplantation recipients and patients with AML evaluated the ability of third-party oral FMT to decrease infection rates[61]. Third-party FMT did not reduce infection rates but was safe and ameliorated intestinal dysbiosis by restoring α-diversity index (and even exceeding baseline values), by restoring commensal bacteria (such as *Collinsella*) and by reducing the abundance of pathobionts (such as *Enterococcus* and *Dialister*). Therefore, both autologous and third-party FMT seem to be safe for AML patients undergoing chemotherapy. In this context, even when a fecal collection at diagnosis before any

antibiotics consumption is possible, the microbial quality of such sample that would be used for autologous FMT remains questionable. Indeed, our work reveals important alterations in the gut microbiota composition and function of AML patients already at diagnosis, which can be linked to metabolic and inflammatory dysregulations in those patients. Our findings call therefore for caution when using autologous fecal material transfer during the therapeutic care of AML patients and goes in favor of heterologous transfer to increase the gut microbiota diversity and richness in these patients.

To our knowledge, no study has been made on FMT and cachexia in AML patients. De Clercq and colleagues[62] performed a double-blind randomized placebo-controlled trial on 24 cachectic patients with metastatic HER2-negative gastroesophageal cancer which received autologous FMT or third-party FMT from a healthy obese donor. Allogenic FMT did not improve any of the cachexia outcomes (such as satiety and caloric intake) but increased disease control rate and showed a tendency of increased overall survival median and progression-free survival. Further research evaluating the impact of FMT on cachexia and its efficacy to tackle it has still to be performed.

## References


1. Ghosh TS, Shanahan F, O'Toole PW. The gut microbiome as a modulator of healthy ageing. Nat Rev Gastroenterol Hepatol. 2022;19(9):565-584.
2. Falony G, Joossens M, Vieira-Silva S, et al. Population-level analysis of gut microbiome variation. Science. 2016;352(6285):560-564.
3. Zhang X, Zhong H, Li Y, et al. Sex- and age-related trajectories of the adult human gut microbiota shared across populations of different ethnicities. Nat Aging. 2021;1(1):87-100.
4. Zhernakova A, Kurilshikov A, Bonder MJ, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352(6285):565-569.
5. Crovesy L, Masterson D, Rosado EL. Profile of the gut microbiota of adults with obesity: a systematic review. Eur J Clin Nutr. 2020;74(9):1251-1262.
6. Antinozzi M, Giffi M, Sini N, et al. Cigarette Smoking and Human Gut Microbiota in Healthy Adults: A Systematic Review. Biomedicines. 2022;10(2):
7. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.
8. . Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research. Washington (DC), 2003.
9. Chua LL, Rajasuriar R, Azanan MS, et al. Reduced microbial diversity in adult survivors of childhood acute lymphoblastic leukemia and microbial associations with increased immune activation. Microbiome. 2017;5(1):35.
10. Hiel S, Bindels LB, Pachikian BD, et al. Effects of a diet based on inulin-rich vegetables on gut health and nutritional behavior in healthy humans. Am J Clin Nutr. 2019;109(6):1683-1695.
11. Leclercq S, Matamoros S, Cani PD, et al. Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. Proc Natl Acad Sci U S A. 2014;111(42):E4485-4493.
12. Dewulf EM, Cani PD, Claus SP, et al. Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women. Gut. 2013;62(8):1112-1121.
13. Nicolucci AC, Hume MP, Martinez I, Mayengbam S, Walter J, Reimer RA. Prebiotics Reduce Body Fat and Alter Intestinal Microbiota in Children Who Are Overweight or With Obesity. Gastroenterology. 2017;153(3):711-722.
14. Rodriguez J, Hiel S, Neyrinck AM, et al. Discovery of the gut microbial signature driving the efficacy of prebiotic intervention in obese patients. Gut. 2020;69(11):1975-1987.
15. Hoebeeck LI, Rietzschel ER, Langlois M, et al. The relationship between diet and subclinical atherosclerosis: results from the Asklepios Study. Eur J Clin Nutr. 2011;65(5):606-613.
16. de l'Homme G, Bacque MF, Housset B, Lebeau B. [Tobacco dependence: a short evaluation questionnaire]. Presse Med. 1992;21(13):606-608.
17. Wilson MM, Thomas DR, Rubenstein LZ, et al. Appetite assessment: simple appetite questionnaire predicts weight loss in community-dwelling adults and nursing home residents. Am J Clin Nutr. 2005;82(5):1074-1081.
18. Holman R, Hines G, Kennedy I, Stevens R, Matthews D, Levy J. A calculator for HOMA. Diabetologia. 2004;47(
19. Costea PI, Zeller G, Sunagawa S, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol. 2017;35(11):1069-1076.
20. Ovreas L, Forney L, Daae FL, Torsvik V. Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. Appl Environ Microbiol. 1997;63(9):3367-3373.
21. Potgens SA, Thibaut MM, Joudiou N, et al. Multi-compartment metabolomics and metagenomics reveal major hepatic and intestinal disturbances in cancer cachectic mice. J Cachexia Sarcopenia Muscle. 2021;12(2):456-475.

22. Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using illumina paired-end technology. PLoS One. 2013;8(6):e66643.
23. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10(10):996-998.
24. Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv. 2016;081257.
25. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-2643.
26. Edgar RC. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ. 2018;6(e4652.
27. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261-5267.
28. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-336.
29. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-7541.
30. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017;8(2224.
31. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Computational Biology. 2017;13(11):
32. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol. 2021;17(11):e1009442.
33. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistic Society: Series B (Methodological). 1995;57(1):289-300.
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-2120.
35. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-359.
36. Beghini F, McIver LJ, Blanco-Miguez A, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife. 2021;10(
37. Oksanen J, Blanchet GF, Friendly M, et al. vegan: community Ecology. 2.5-7 ed: R package, 2020.
38. Nearing JT, Douglas GM, Hayes MG, et al. Microbiome differential abundance methods produce different results across 38 datasets. Nature Communications. 2022;13(1):342.
39. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2(15.
40. Podlesny D, Durdevic M, Paramsothy S, et al. Identification of clinical and ecological determinants of strain engraftment after fecal microbiota transplantation using metagenomics. Cell Rep Med. 2022;3(8):100711.
41. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. mSystems. 2018;3(6):
42. Tuszynski J, Dietze M. Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc. https://cran.r-project.org/src/contrib/Archive/caTools/, 2005.
43. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2(3):18-22.
44. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. http://rocr.bioinf.mpi-sb.mpg.de. Bioinformatics. 2005;21(20):

45. Sousa SAA, Magalhães A, Ferreira MMC. Optimized bucketing for NMR spectra: Three case studies. Chemometrics and Intelligent Laboratory Systems. 2013;122(93-102.
46. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 2018;46(D1):D608-D617.
47. Lazar C. imputeLCMD: a collection of methods for left-censored missing data imputation. CRAN, R Package, version 20. 2015;
48. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4(1686.
49. Janssens Y, Nielandt J, Bronselaer A, et al. Disbiome database: linking the microbiome to disease. BMC Microbiol. 2018;18(1):50.
50. Qi C, Cai Y, Qian K, et al. gutMDisorder v2.0: a comprehensive database for dysbiosis of gut microbiota in phenotypes and interventions. Nucleic Acids Res. 2023;51(D1):D717-D722.
51. Brahe LK, Le Chatelier E, Prifti E, et al. Specific gut microbiota features and metabolic markers in postmenopausal women with obesity. Nutr Diabetes. 2015;5(6):e159.
52. de Goffau MC, Luopajarvi K, Knip M, et al. Fecal microbiota composition differs between children with beta-cell autoimmunity and those without. Diabetes. 2013;62(4):1238-1244.
53. Ignacio A, Fernandes MR, Rodrigues VA, et al. Correlation between body mass index and faecal microbiota from children. Clin Microbiol Infect. 2016;22(3):258 e251-258.
54. Giongo A, Gano KA, Crabb DB, et al. Toward defining the autoimmune microbiome for type 1 diabetes. ISME J. 2011;5(1):82-91.
55. Maya-Lucas O, Murugesan S, Nirmalkar K, et al. The gut microbiome of Mexican children affected by obesity. Anaerobe. 2019;55(11-23.
56. D'Angelo CR, Sudakaran S, Callander NS. Clinical effects and applications of the gut microbiome in hematologic malignancies. Cancer. 2021;127(5):679-687.
57. Galloway-Pena JR, Smith DP, Sahasrabhojane P, et al. The role of the gastrointestinal microbiome in infectious complications during induction chemotherapy for acute myeloid leukemia. Cancer. 2016;122(14):2186-2196.
58. Rashidi A, Ebadi M, Rehman TU, et al. Lasting shift in the gut microbiota in patients with acute myeloid leukemia. Blood Adv. 2022;6(11):3451-3457.
59. Rattanathammethee T, Tuitemwong P, Thiennimitr P, et al. Gut microbiota profiles of treatment-naive adult acute myeloid leukemia patients with neutropenic fever during intensive chemotherapy. PLoS One. 2020;15(10):e0236460.
60. Malard F, Vekhoff A, Lapusan S, et al. Gut microbiota diversity after autologous fecal microbiota transfer in acute myeloid leukemia patients. Nat Commun. 2021;12(1):3084.
61. Rashidi A, Ebadi M, Rehman TU, et al. Randomized Double-Blind Phase II Trial of Fecal Microbiota Transplantation Versus Placebo in Allogeneic Hematopoietic Cell Transplantation and AML. J Clin Oncol. 2023;41(34):5306-5319.
62. de Clercq NC, van den Ende T, Prodan A, et al. Fecal Microbiota Transplantation from Overweight or Obese Donors in Cachectic Patients with Advanced Gastroesophageal Cancer: A Randomized, Double-blind, Placebo-Controlled, Phase II Study. Clin Cancer Res. 2021;27(13):3784-3792.

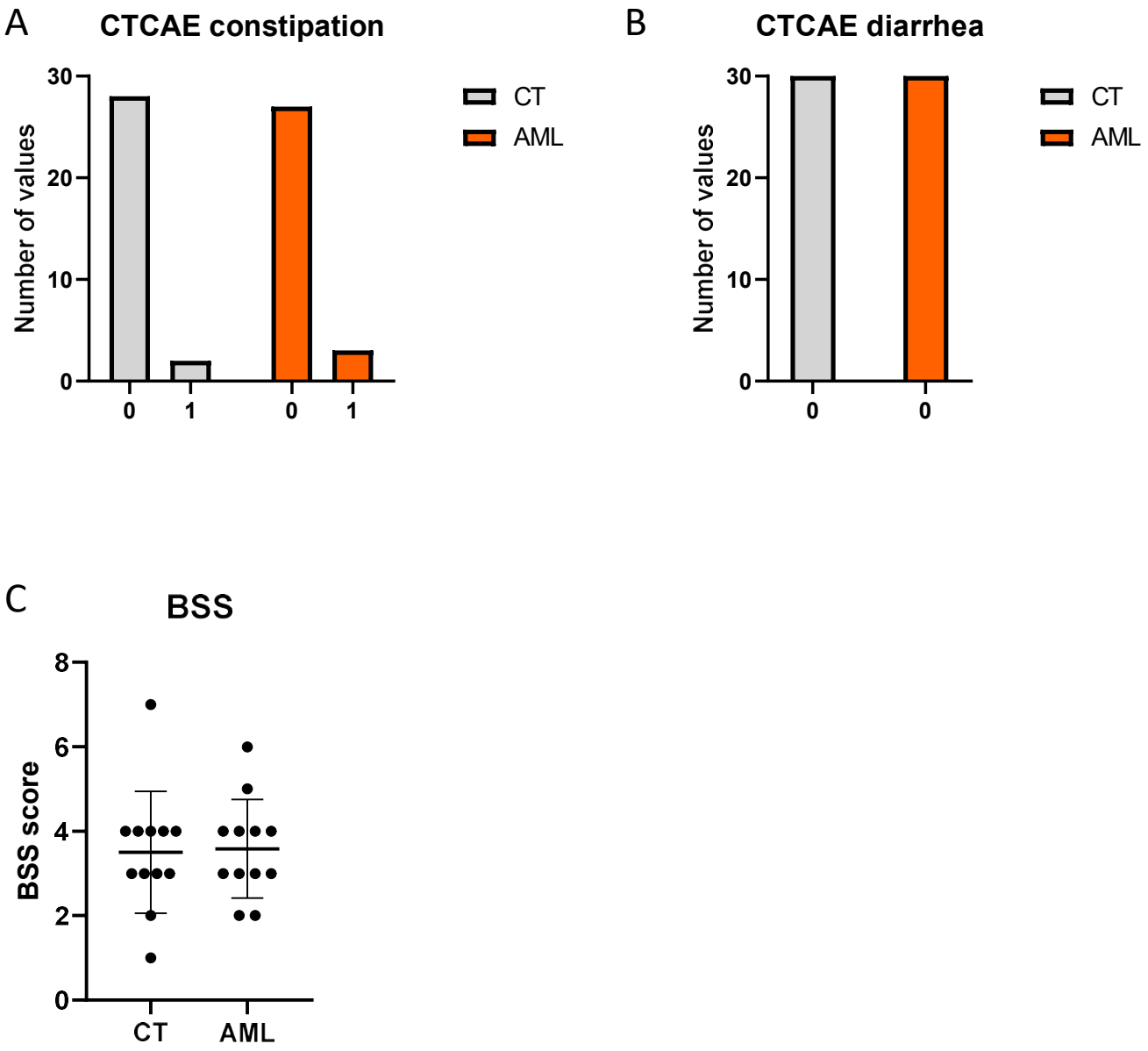



**Figure S1. AML patients do not display transit alterations compared to CT subjects**.
A) CTCAE constipation scores. B) CTCAE diarrhea scores. C) Bristol stool scale (BSS) scores for 12 matched AML and CT subjects. AML in orange *vs.* CT in grey.

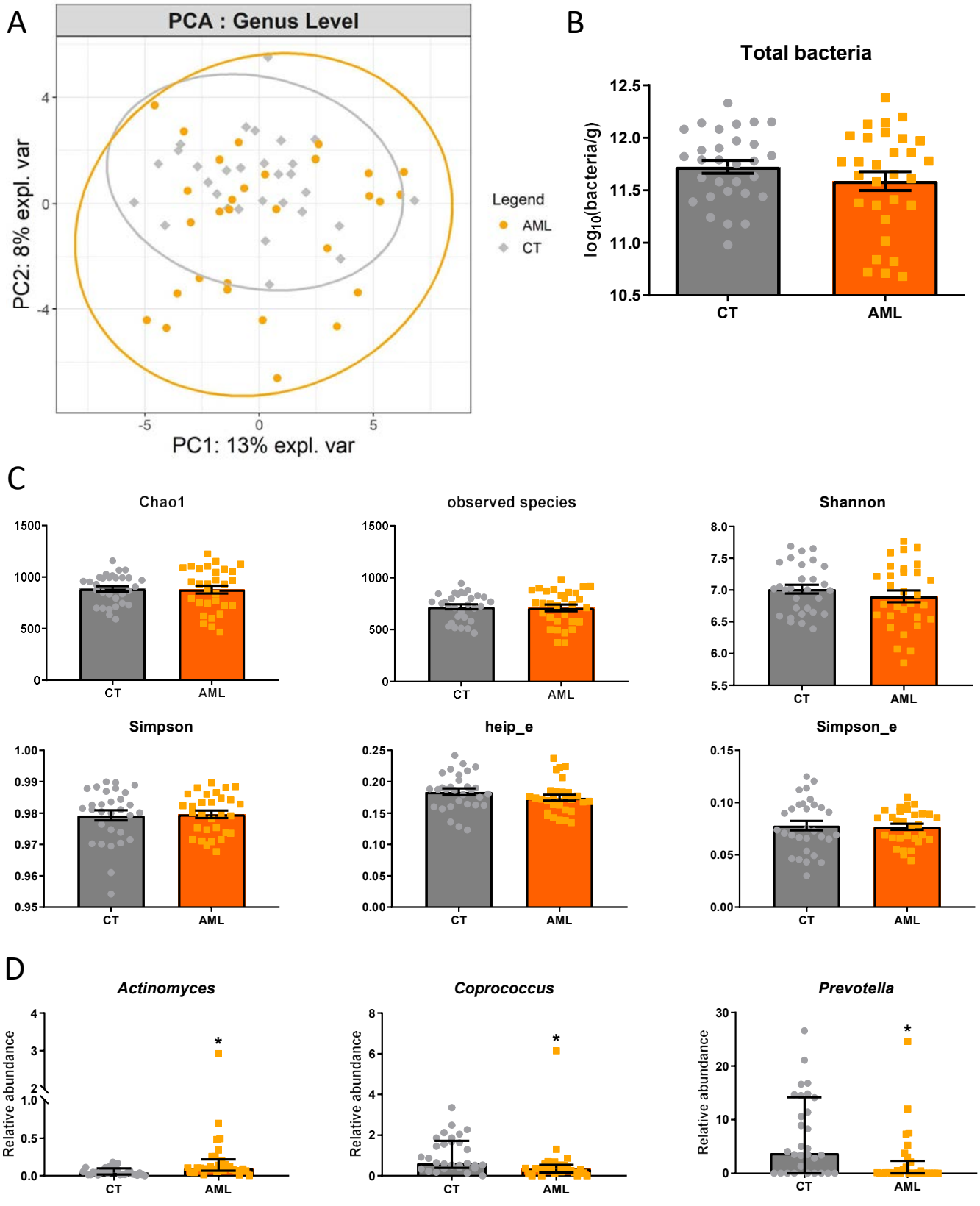



**Figure S2. Alterations in the gut microbiota composition in CT subjects and AML patients (results of 16S rRNA gene sequencing).**

A) Principal component analysis (PCA) at the genus level. PERMANOVA: $R^2$ = 2.7% * B) Total bacteria levels measured by qPCR. C) α-diversity indexes. Indexes that are normally distributed are expressed as mean (standard deviation) and are tested using a Student t-test or a Welch's t test. Indexes that are non-normally distributed are expressed as median (interquartile range) and are tested by a Mann-Whitney U-test. D) Significant changes at the lowest taxa level. Mann-Whitney U-tests with an FDR correction were applied. All q-values < 0.1. *n* = 30. AML in orange *vs.* CT in grey. *: p-value < 0.05

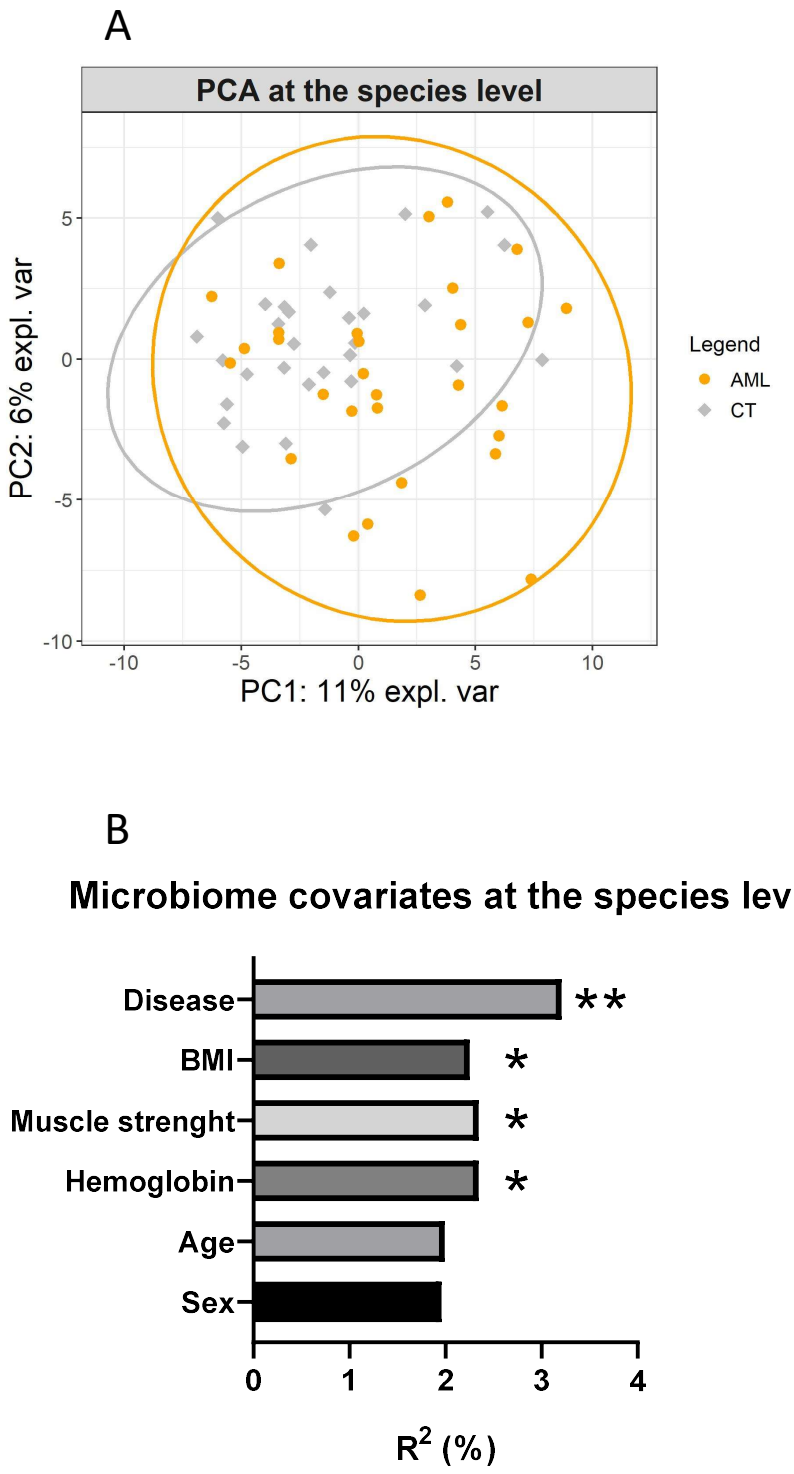
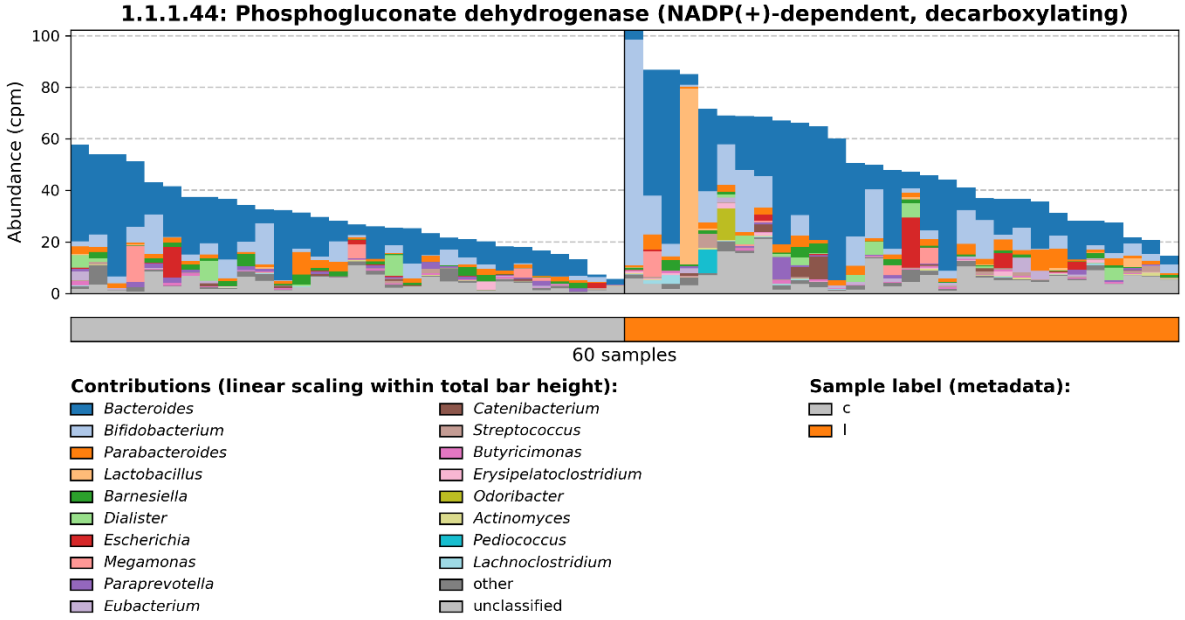
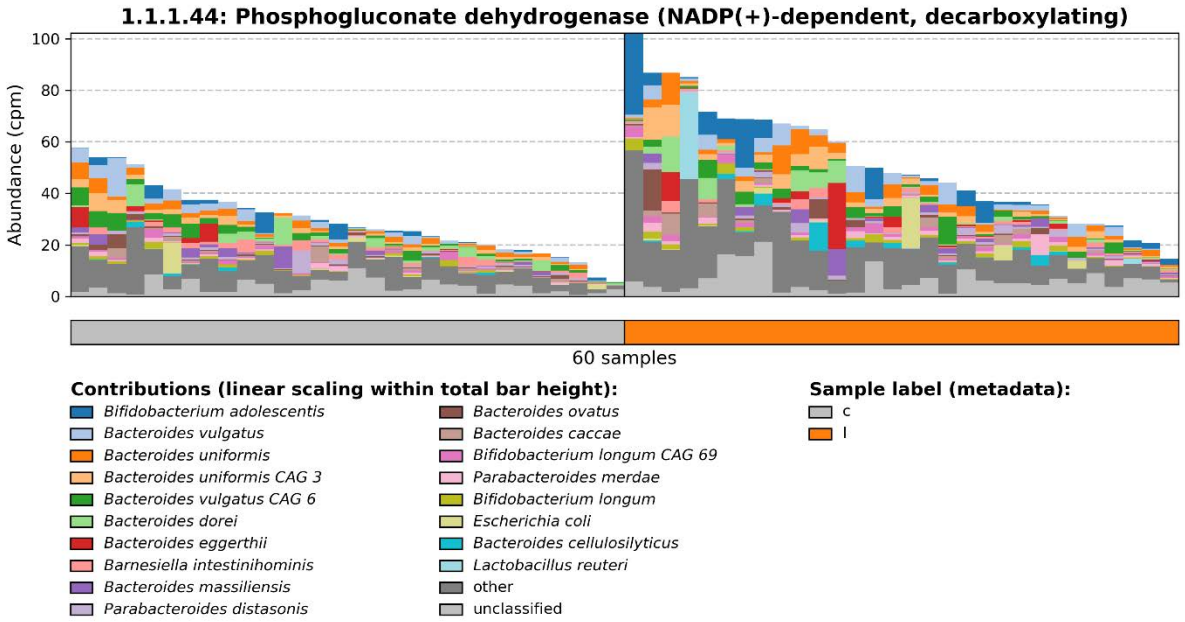


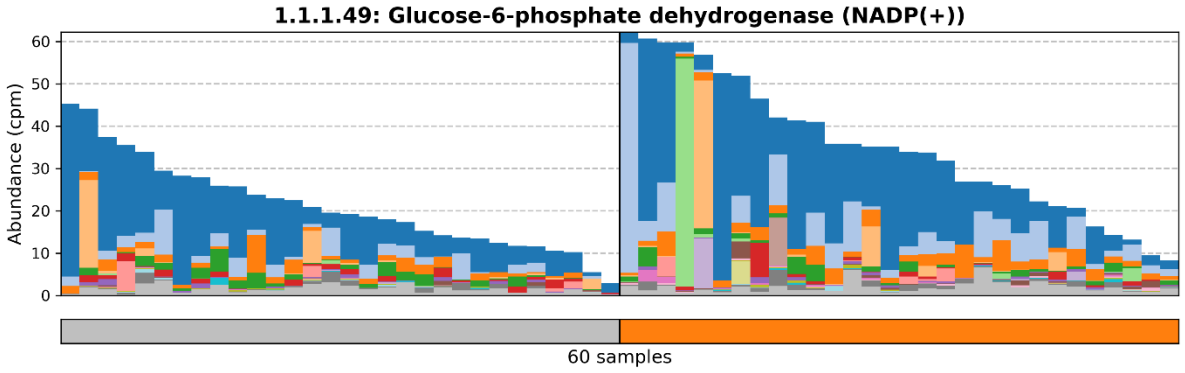

**Figure S3. Alterations in the gut microbiota composition in CT subjects and AML patients (results of metagenomics sequencing).**

A) Principal component analysis (PCA) at the species level. PERMANOVA: $R^2$ = 3.2% **. B) Contribution of disease, BMI, muscle strength, hemoglobin, age and sex to the variance in the PCA at the species level (PERMANOVA results). **p-value < 0.01; *p-value < 0.05

**Figure S4. Bacterial contribution, per genera and per species, to the bacterial functions determined by metagenomics, that are significantly changed between CT (c) and AML (l).** The relative abundance of each function is presented in Fig 1. Cpm, count per million. Plots were drawn using the *human_barplot* function in HUMAnN 3.00.



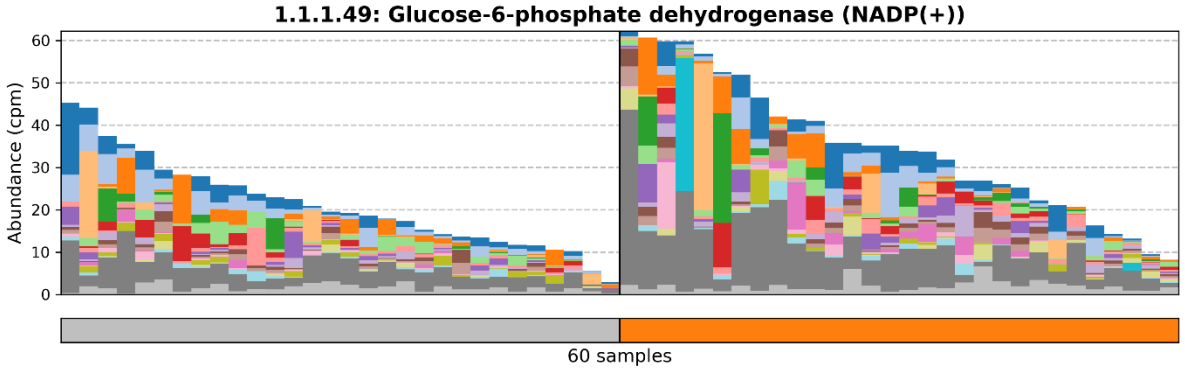



**Contributions (linear scaling within total bar height):**

- *Bacteroides*
- *Bifidobacterium*
- *Parabacteroides*
- *Escherichia*
- *Barnesiella*
- *Lactobacillus*
- *Paraprevotella*
- *Megamonas*
- *Butyricimonas*
- *Klebsiella*
- *Streptococcus*
- *Odoribacter*
- *Lachnoclostridium*
- *Actinomyces*
- *Coprobacter*
- *Pediococcus*
- *Alistipes*
- *Haemophilus*
- other
- unclassified

**Sample label (metadata):**

- c
- l

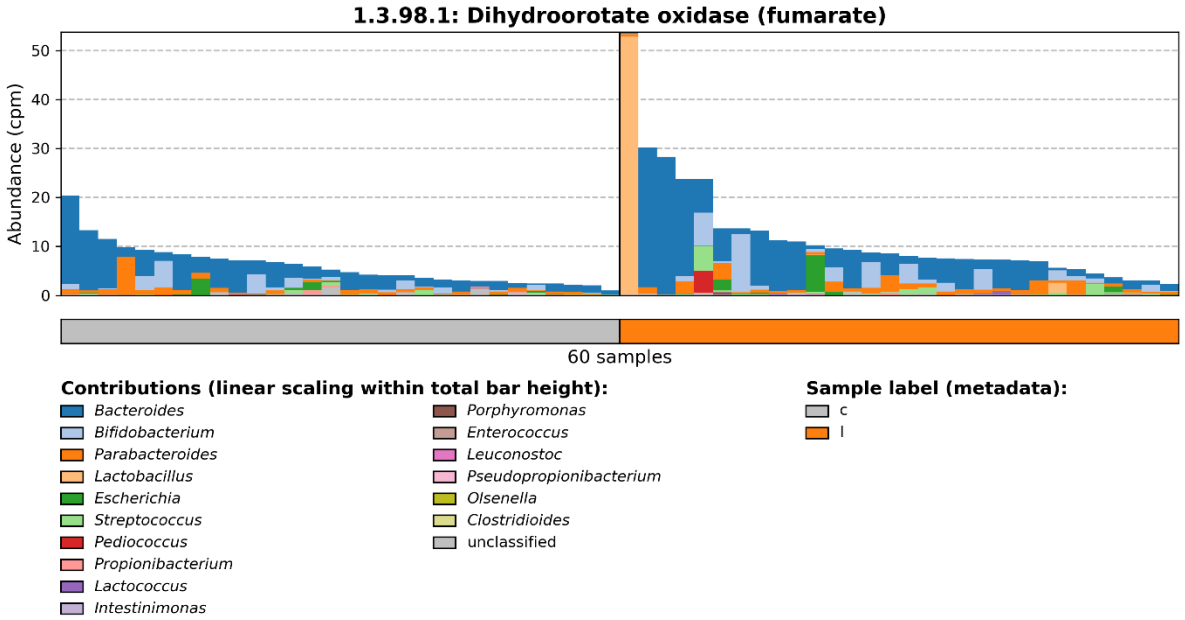


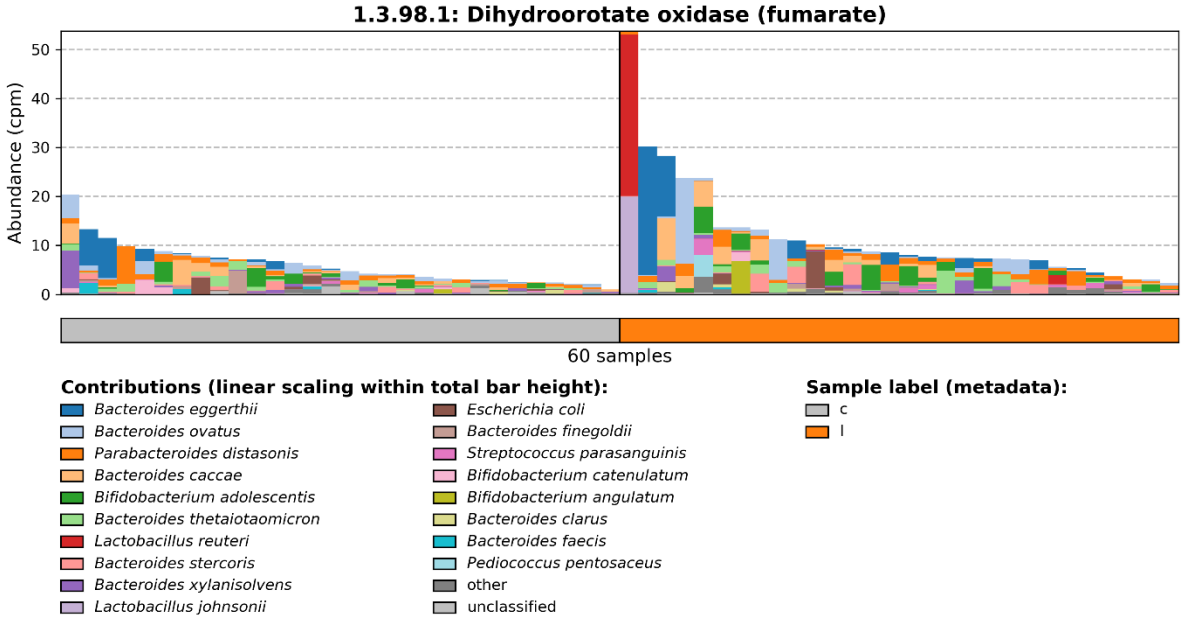

**Contributions (linear scaling within total bar height):**

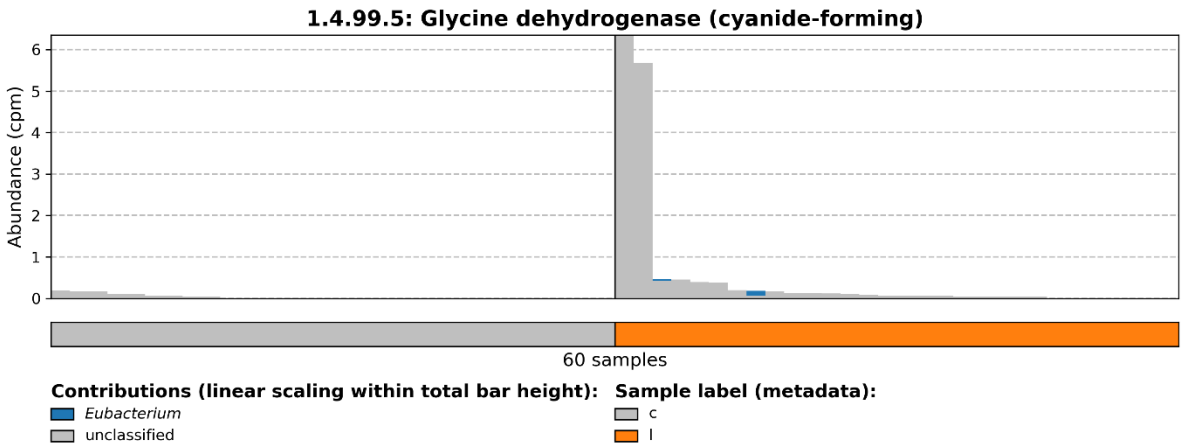
- *Bacteroides vulgatus*
- *Bacteroides vulgatus CAG 6*
- *Bacteroides dorei*
- *Escherichia coli*
- *Bacteroides eggerthii*
- *Barnesiella intestinihominis*
- *Bacteroides massiliensis*
- *Parabacteroides distasonis*
- *Bacteroides caccae*
- *Parabacteroides merdae*
- *Bifidobacterium longum*
- *Bifidobacterium longum CAG 69*
- *Bacteroides cellulosilyticus*
- *Bacteroides ovatus*
- *Paraprevotella xylaniphila*
- *Bifidobacterium adolescentis*
- *Lactobacillus reuteri*
- *Bacteroides thetaiotaomicron*
- other
- unclassified

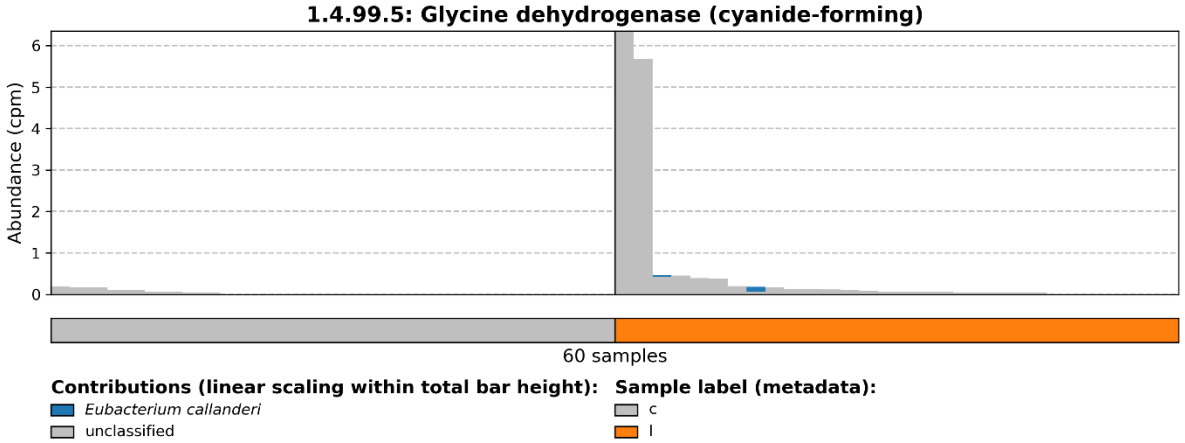
**Sample label (metadata):**

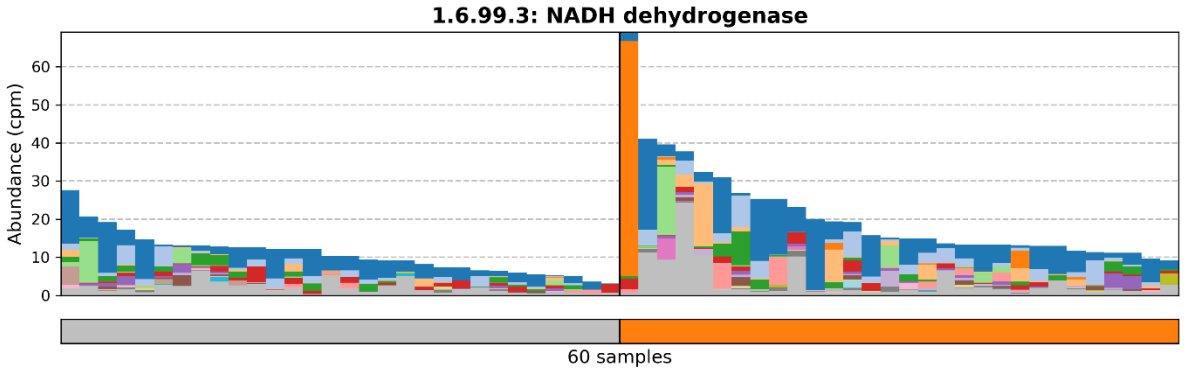
- c
- l


1.3.98.1: Dihydroorotate oxidase (fumarate)
Abundance (cpm)
0
10
20
30
40
50
60 samples
Contributions (linear scaling within total bar height):
Bacteroides
Bifidobacterium
Parabacteroides
Lactobacillus
Escherichia
Streptococcus
Pediococcus
Propionibacterium
Lactococcus
Intestinimonas
Porphyromonas
Enterococcus
Leuconostoc
Pseudopropionibacterium
Olsenella
Clostridioides
unclassified
Sample label (metadata):
c
l


1.3.98.1: Dihydroorotate oxidase (fumarate)
Abundance (cpm)
0
10
20
30
40
50
60 samples
Contributions (linear scaling within total bar height):
Bacteroides eggerthii
Bacteroides ovatus
Parabacteroides distasonis
Bacteroides caccae
Bifidobacterium adolescentis
Bacteroides thetaiotaomicron
Lactobacillus reuteri
Bacteroides stercoris
Bacteroides xylanisolvens
Lactobacillus johnsonii
Escherichia coli
Bacteroides finegoldii
Streptococcus parasanguinis
Bifidobacterium catenulatum
Bifidobacterium angulatum
Bacteroides clarus
Bacteroides faecis
Pediococcus pentosaceus
other
unclassified
Sample label (metadata):
c
l


1.4.99.5: Glycine dehydrogenase (cyanide-forming)
Abundance (cpm)
0
1
2
3
4
5
6
60 samples
Contributions (linear scaling within total bar height):
Eubacterium
unclassified
Sample label (metadata):
c
l


1.4.99.5: Glycine dehydrogenase (cyanide-forming)
Abundance (cpm)
0
1
2
3
4
5
6
60 samples
Contributions (linear scaling within total bar height):
Eubacterium callanderi
unclassified
Sample label (metadata):
c
l

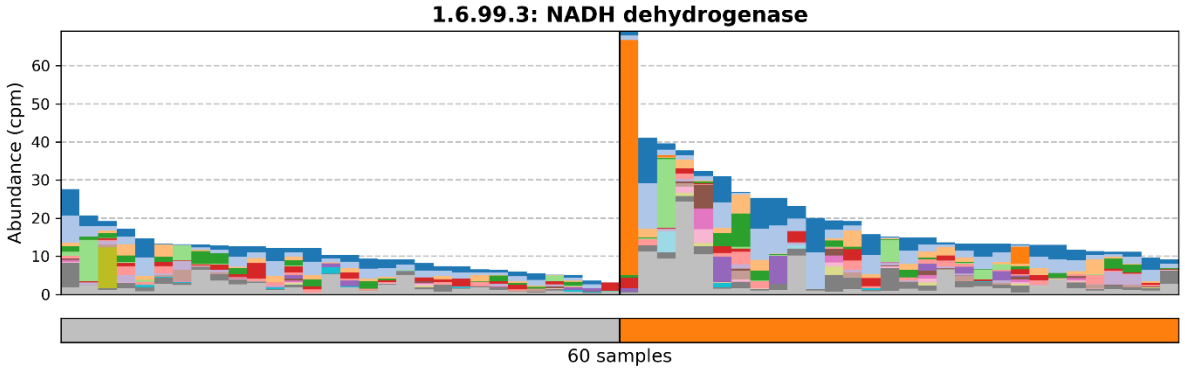



**Contributions (linear scaling within total bar height):**

- *Bacteroides*
- *Bifidobacterium*
- *Lactobacillus*
- *Streptococcus*
- *Fusicatenibacter*
- *Escherichia*
- *Coprococcus*
- *Catenibacterium*
- *Clostridium*
- *Actinomyces*
- *Romboutsia*
- *Hafnia*
- *Klebsiella*
- *Haemophilus*
- *Lactococcus*
- *Blautia*
- *Enterococcus*
- *Hungatella*
- other
- unclassified

**Sample label (metadata):**

- c
- l

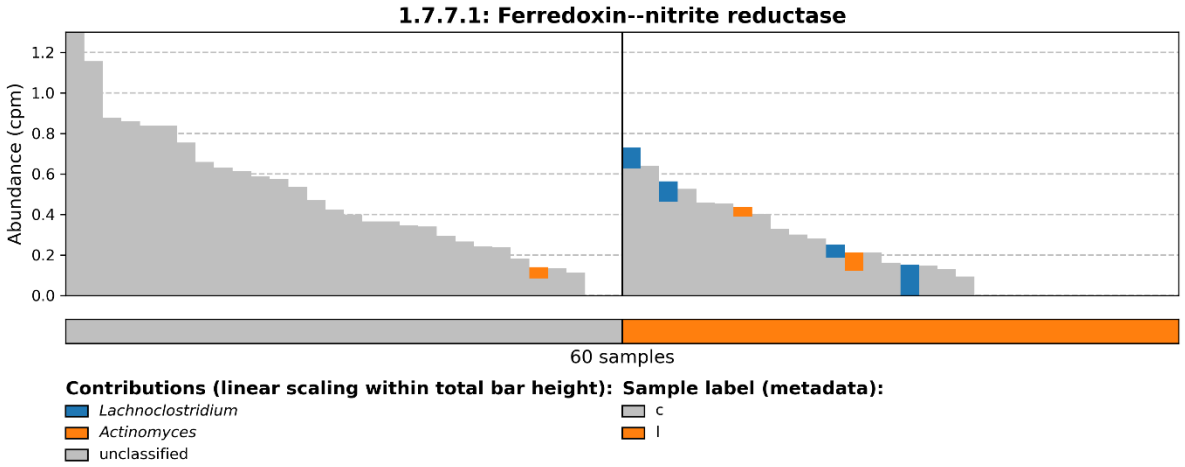


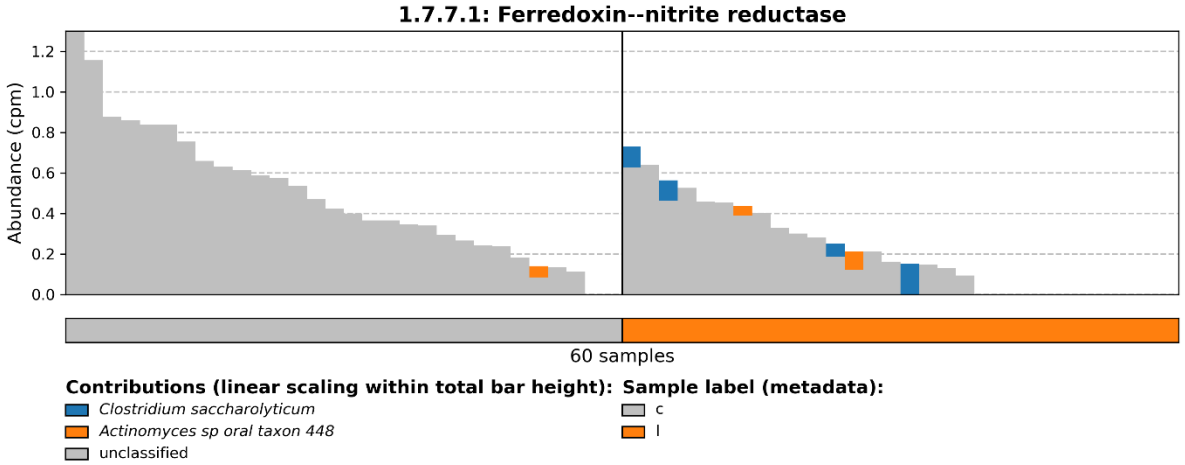

**Contributions (linear scaling within total bar height):**

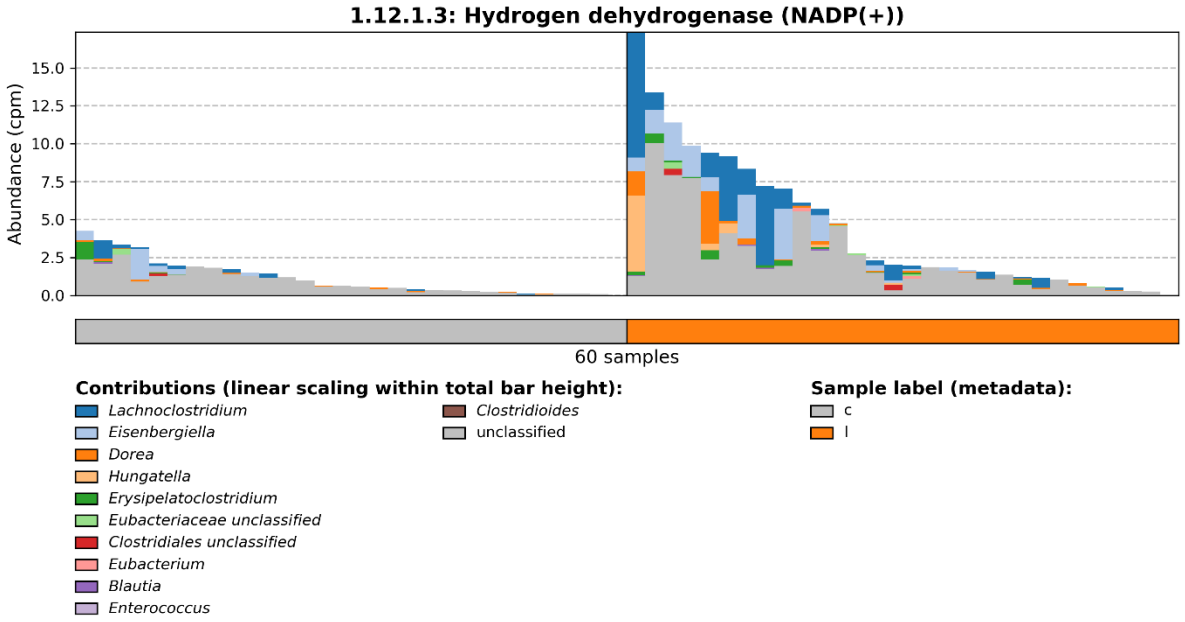
- *Bacteroides uniformis*
- *Bacteroides uniformis CAG 3*
- *Lactobacillus reuteri*
- *Bifidobacterium longum CAG 69*
- *Fusicatenibacter saccharivorans*
- *Escherichia coli*
- *Coprococcus comes CAG 19*
- *Bifidobacterium longum*
- *Catenibacterium mitsuokai*
- *Clostridium disporicum*
- *Streptococcus salivarius*
- *Romboutsia ilealis*
- *Streptococcus thermophilus*
- *Streptococcus salivarius CAG 79*
- *Bacteroides xylanisolvens*
- *Streptococcus parasanguinis*
- *Bacteroides finegoldii*
- *Klebsiella pneumoniae*
- other
- unclassified

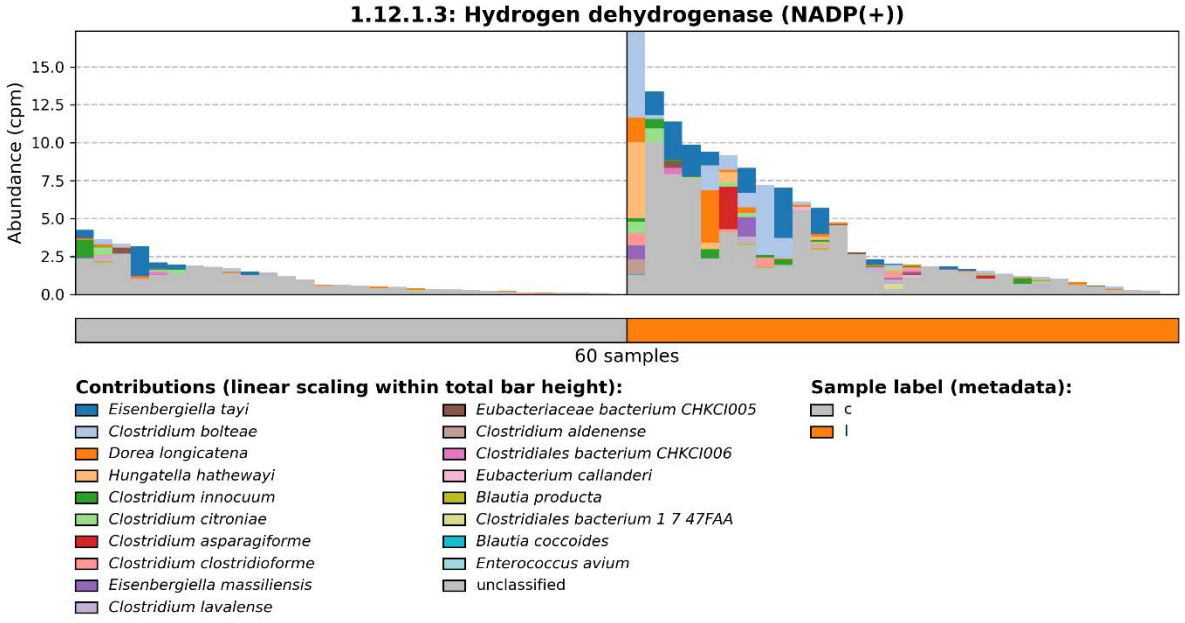
**Sample label (metadata):**

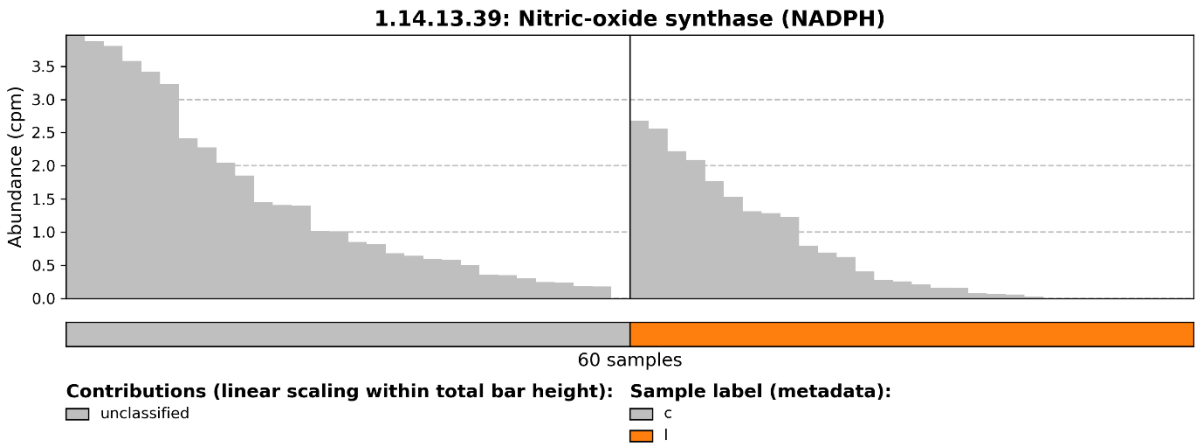
- c
- l

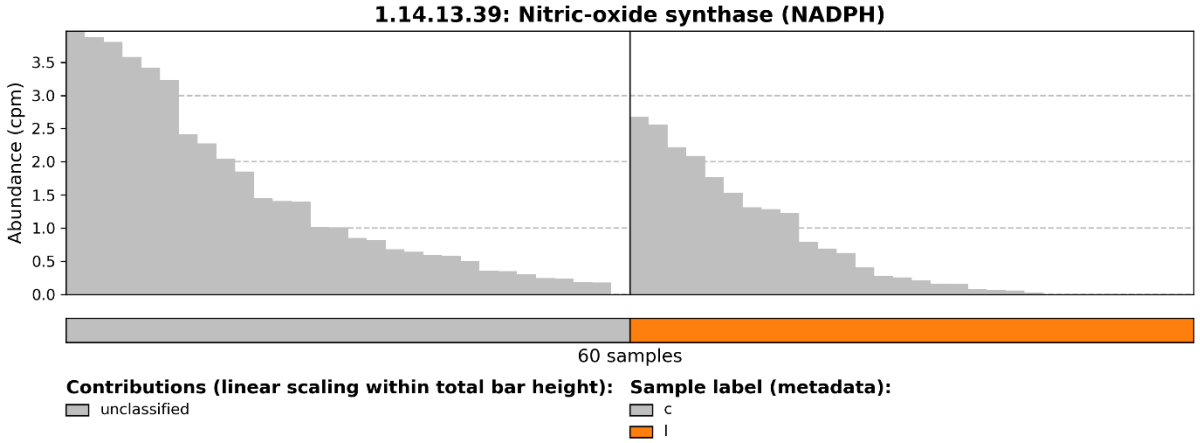

1.7.7.1: Ferredoxin--nitrite reductase
Abundance (cpm)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
60 samples
Contributions (linear scaling within total bar height):
Lachnoclostridium
Actinomyces
unclassified
Sample label (metadata):
c
l

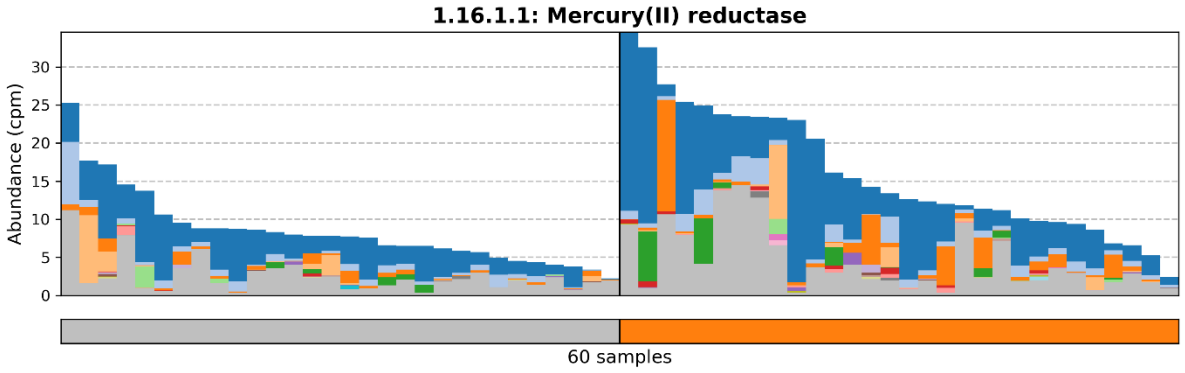

1.7.7.1: Ferredoxin--nitrite reductase
Abundance (cpm)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
60 samples
Contributions (linear scaling within total bar height):
Clostridium saccharolyticum
Actinomyces sp oral taxon 448
unclassified
Sample label (metadata):
c
l


1.12.1.3: Hydrogen dehydrogenase (NADP(+))
Abundance (cpm)
15.0
12.5
10.0
7.5
5.0
2.5
0.0
60 samples
Contributions (linear scaling within total bar height):
Lachnoclostridium
Eisenbergiella
Dorea
Hungatella
Erysipelatoclostridium
Eubacteriaceae unclassified
Clostridiales unclassified
Eubacterium
Blautia
Enterococcus
Clostridioides
unclassified
Sample label (metadata):
c
l


1.12.1.3: Hydrogen dehydrogenase (NADP(+))
Abundance (cpm)
15.0
12.5
10.0
7.5
5.0
2.5
0.0
60 samples
Contributions (linear scaling within total bar height):
Eisenbergiella tayi
Clostridium bolteae
Dorea longicatena
Hungatella hathewayi
Clostridium innocuum
Clostridium citroniae
Clostridium asparagiforme
Clostridium clostridioforme
Eisenbergiella massiliensis
Clostridium lavalense
Eubacteriaceae bacterium CHKCI005
Clostridium aldenense
Clostridiales bacterium CHKCI006
Eubacterium callanderi
Blautia producta
Clostridiales bacterium 1 7 47FAA
Blautia coccoides
Enterococcus avium
unclassified
Sample label (metadata):
c
l


1.14.13.39: Nitric-oxide synthase (NADPH)
Abundance (cpm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
60 samples
Contributions (linear scaling within total bar height):
unclassified
Sample label (metadata):
c
l


1.14.13.39: Nitric-oxide synthase (NADPH)
Abundance (cpm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
60 samples
Contributions (linear scaling within total bar height):
unclassified
Sample label (metadata):
c
l

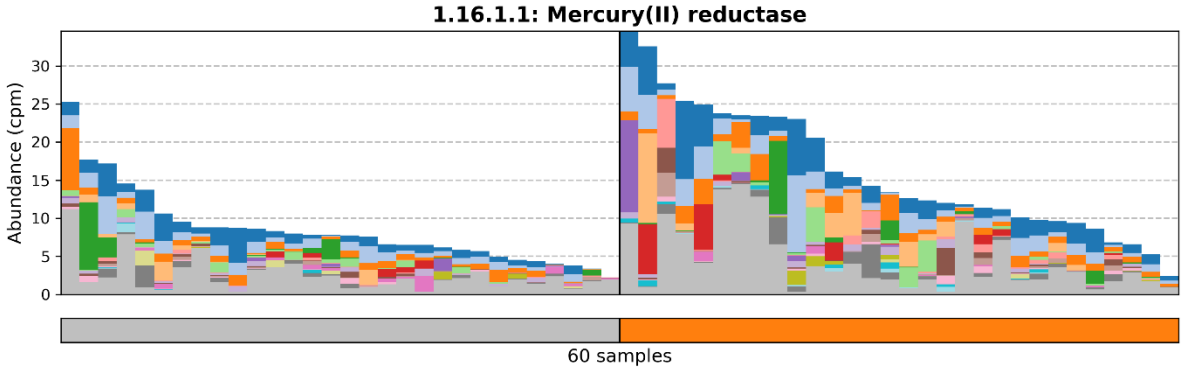



**Contributions (linear scaling within total bar height):**

- *Bacteroides*
- *Parabacteroides*
- *Streptococcus*
- *Escherichia*
- *Catenibacterium*
- *Lactobacillus*
- *Eisenbergiella*
- *Erysipelatoclostridium*
- *Lactococcus*
- *Rothia*
- *Haemophilus*
- *Actinomyces*
- *Clostridium*
- *Bifidobacterium*
- *Lachnoclostridium*
- *Veillonella*
- *Enterococcus*
- *Anaeromassilibacillus*
- other
- unclassified

**Sample label (metadata):**

- c
- l

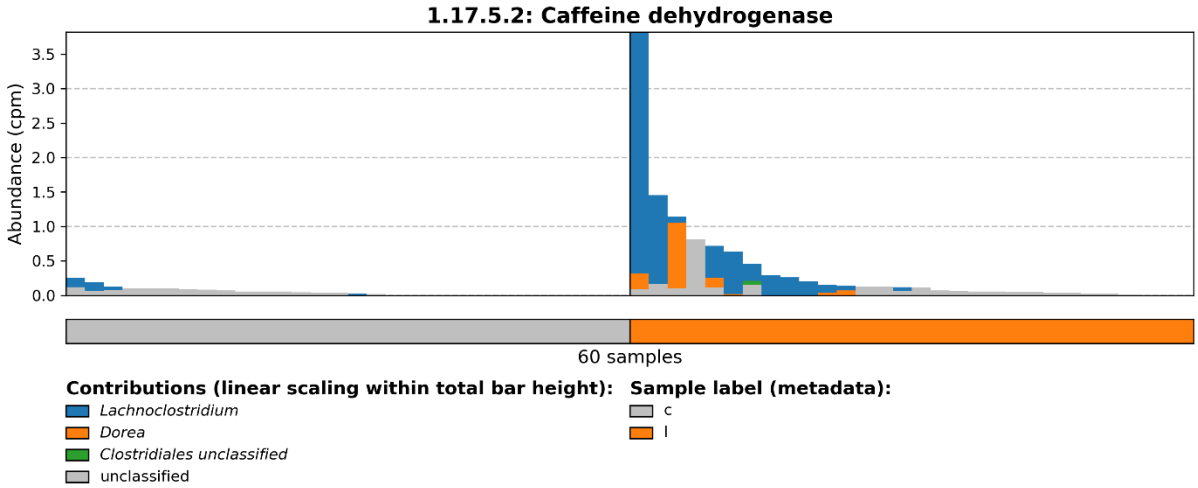


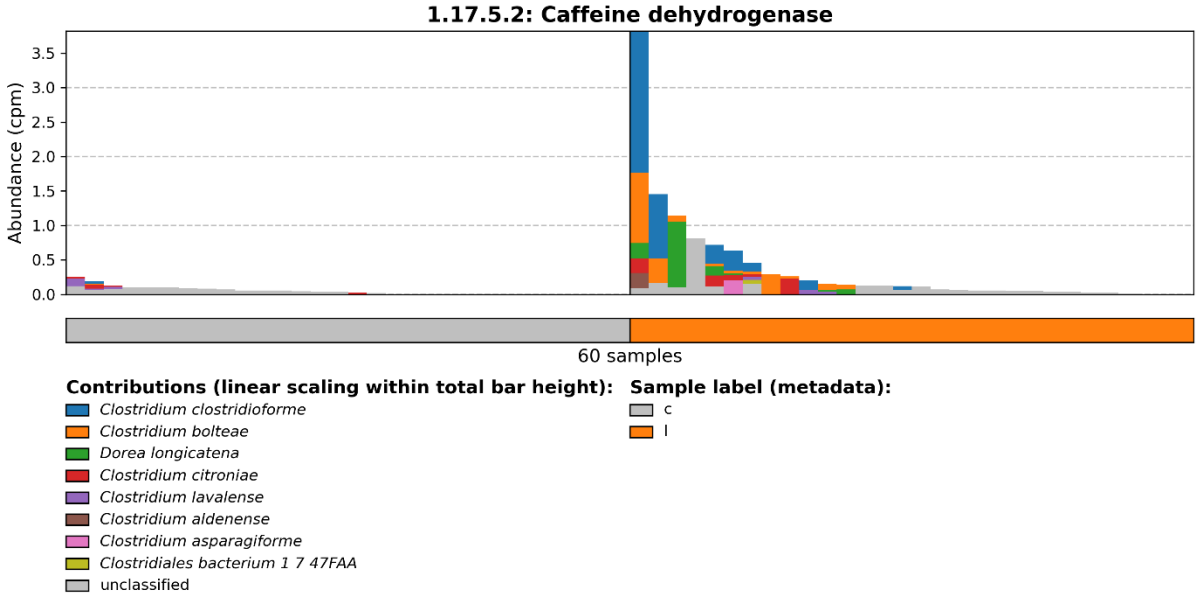

**Contributions (linear scaling within total bar height):**

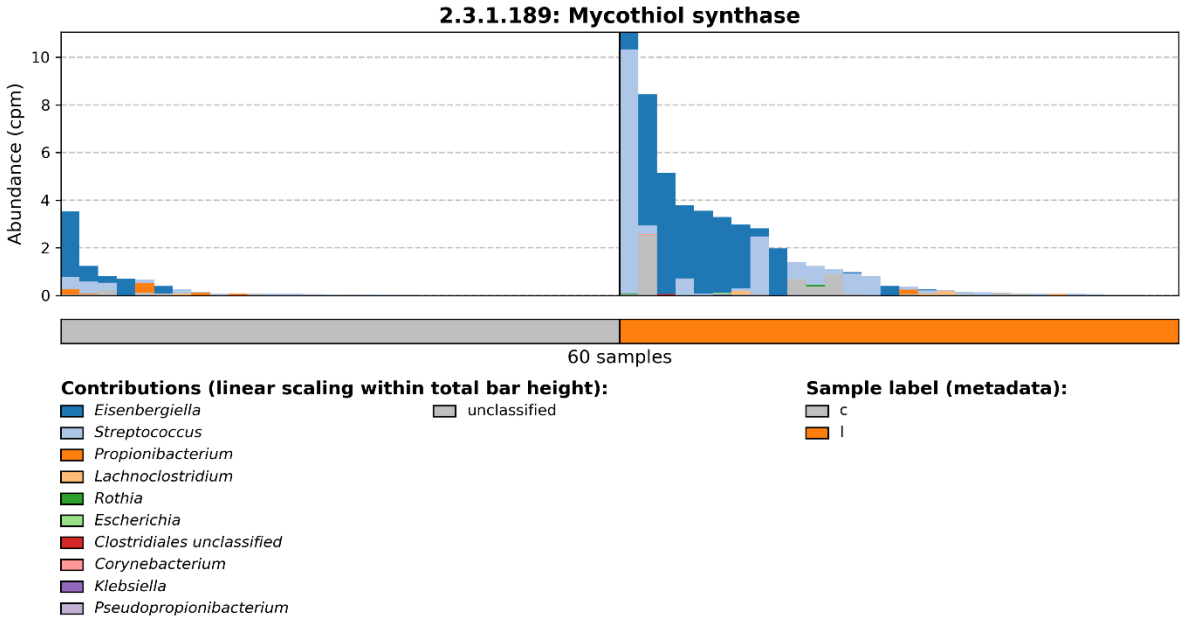
- *Bacteroides uniformis CAG 3*
- *Bacteroides uniformis*
- *Parabacteroides distasonis*
- *Bacteroides cellulosilyticus*
- *Escherichia coli*
- *Bacteroides stercoris*
- *Catenibacterium mitsuokai*
- *Streptococcus salivarius*
- *Bacteroides intestinalis*
- *Bacteroides thetaiotaomicron*
- *Streptococcus thermophilus CAG 236*
- *Streptococcus salivarius CAG 79*
- *Bacteroides finegoldii*
- *Streptococcus thermophilus*
- *Bacteroides clarus*
- *Bacteroides faecis*
- *Eisenbergiella tayi*
- *Clostridium innocuum*
- other
- unclassified

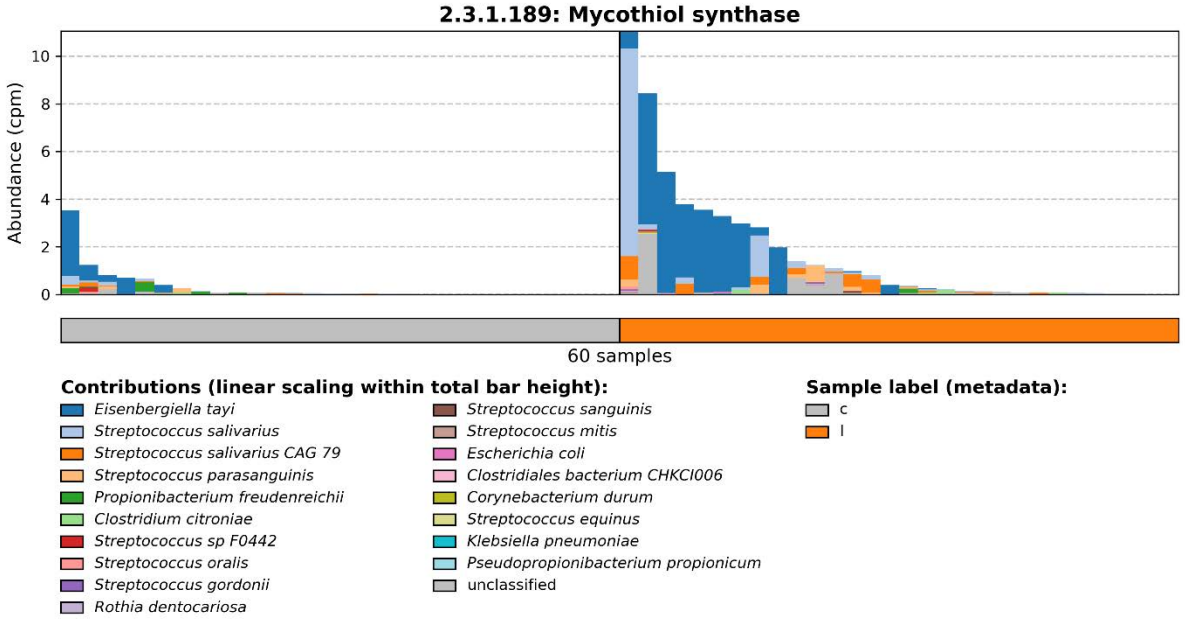
**Sample label (metadata):**

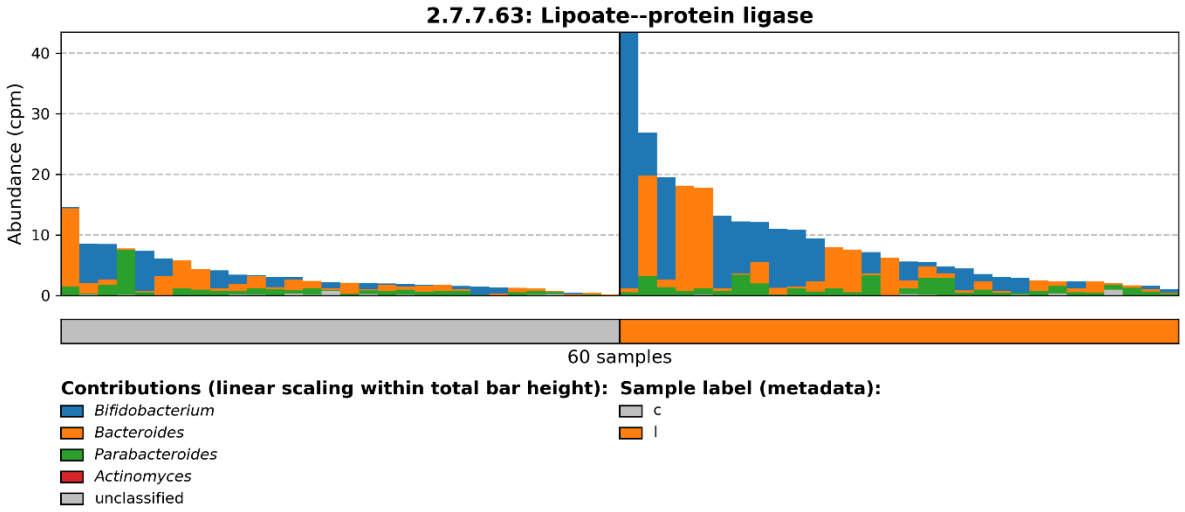
- c
- l

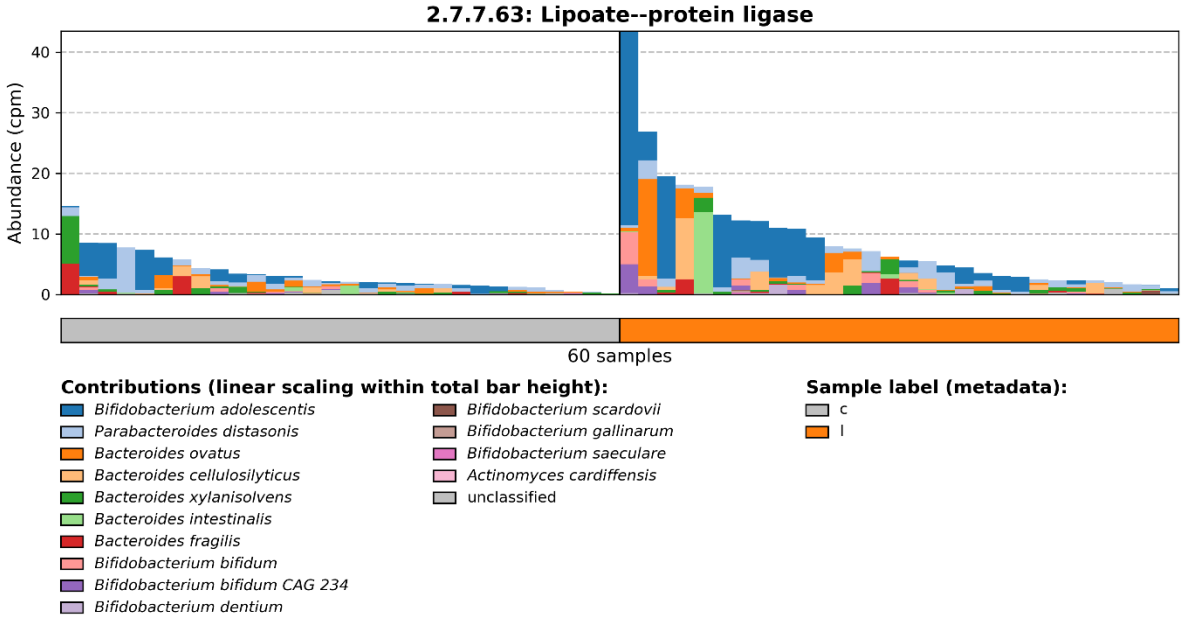

1.17.5.2: Caffeine dehydrogenase
Abundance (cpm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
60 samples
Contributions (linear scaling within total bar height):
Lachnoclostridium
Dorea
Clostridiales unclassified
unclassified
Sample label (metadata):
c
l

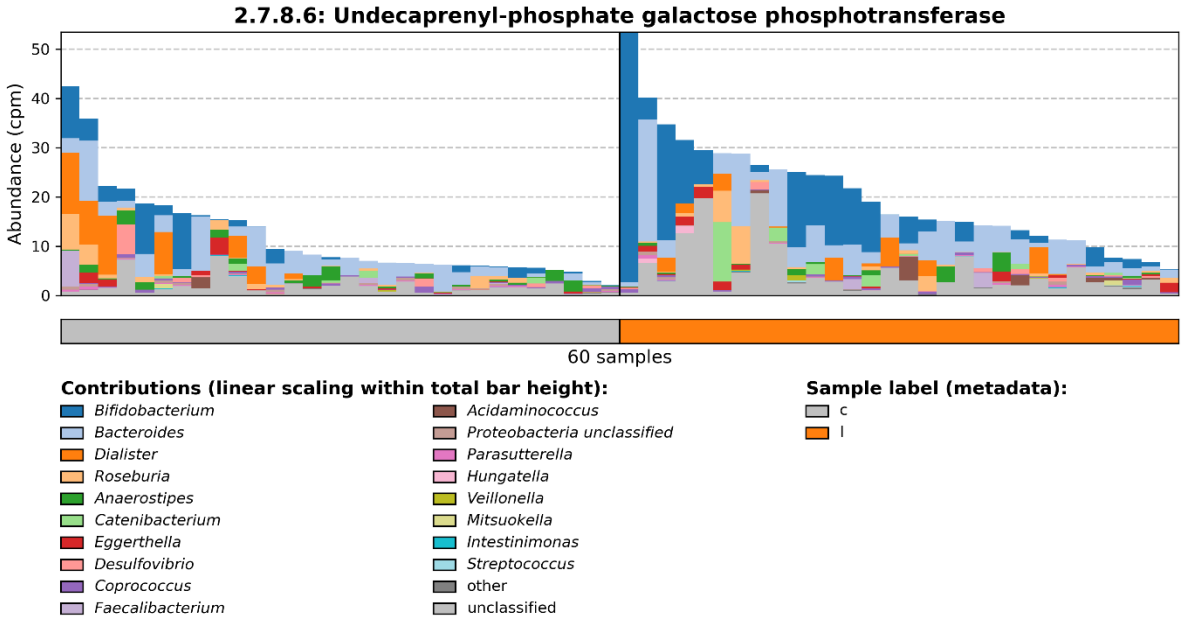

1.17.5.2: Caffeine dehydrogenase
Abundance (cpm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
60 samples
Contributions (linear scaling within total bar height):
Clostridium clostridioforme
Clostridium bolteae
Dorea longicatena
Clostridium citroniae
Clostridium lavalense
Clostridium aldenense
Clostridium asparagiforme
Clostridiales bacterium 1 7 47FAA
unclassified
Sample label (metadata):
c
l

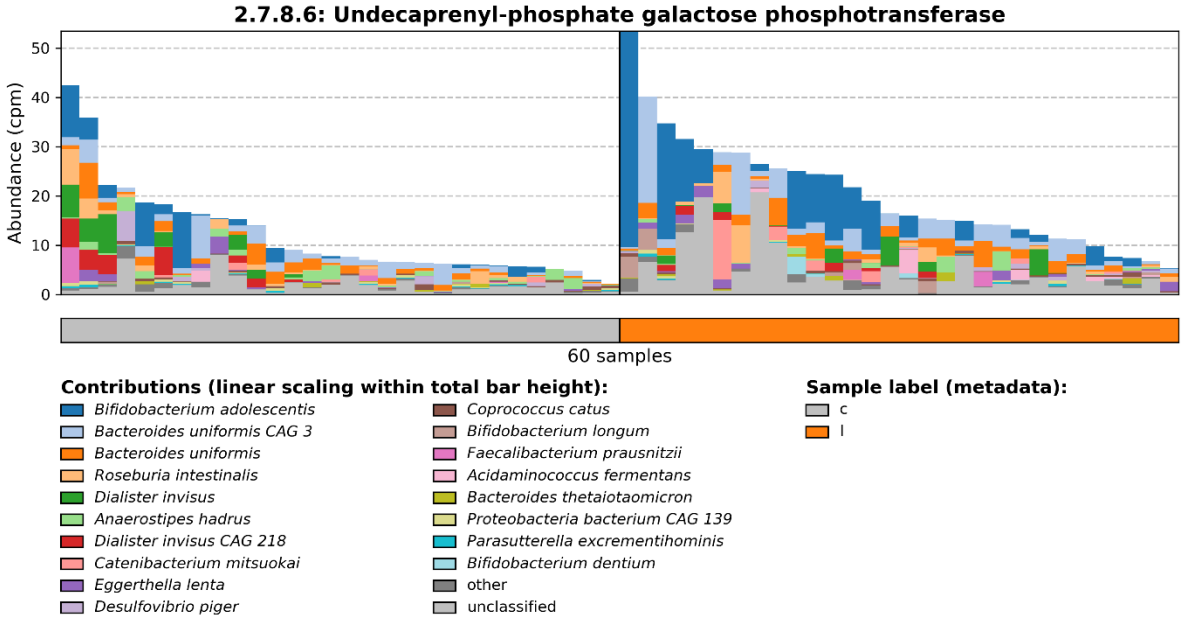

2.3.1.189: Mycothiol synthase
Abundance (cpm)
0
2
4
6
8
10
60 samples
Contributions (linear scaling within total bar height):
Eisenbergiella
Streptococcus
Propionibacterium
Lachnoclostridium
Rothia
Escherichia
Clostridiales unclassified
Corynebacterium
Klebsiella
Pseudopropionibacterium
unclassified
Sample label (metadata):
c
l

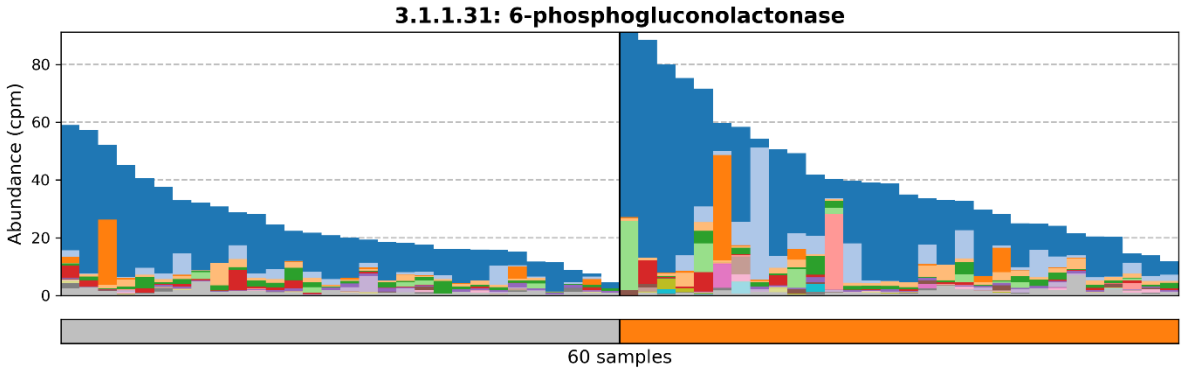

2.3.1.189: Mycothiol synthase
Abundance (cpm)
0
2
4
6
8
10
60 samples
Contributions (linear scaling within total bar height):
Eisenbergiella tayi
Streptococcus salivarius
Streptococcus salivarius CAG 79
Streptococcus parasanguinis
Propionibacterium freudenreichii
Clostridium citroniae
Streptococcus sp F0442
Streptococcus oralis
Streptococcus gordonii
Rothia dentocariosa
Streptococcus sanguinis
Streptococcus mitis
Escherichia coli
Clostridiales bacterium CHKCI006
Corynebacterium durum
Streptococcus equinus
Klebsiella pneumoniae
Pseudopropionibacterium propionicum
unclassified
Sample label (metadata):
c
l


2.7.7.63: Lipoate--protein ligase
Abundance (cpm)
0
10
20
30
40
60 samples
Contributions (linear scaling within total bar height):
Bifidobacterium
Bacteroides
Parabacteroides
Actinomyces
unclassified
Sample label (metadata):
c
l


2.7.7.63: Lipoate--protein ligase
Abundance (cpm)
0
10
20
30
40
60 samples
Contributions (linear scaling within total bar height):
Bifidobacterium adolescentis
Parabacteroides distasonis
Bacteroides ovatus
Bacteroides cellulosilyticus
Bacteroides xylanisolvens
Bacteroides intestinalis
Bacteroides fragilis
Bifidobacterium bifidum
Bifidobacterium bifidum CAG 234
Bifidobacterium dentium
Bifidobacterium scardovii
Bifidobacterium gallinarum
Bifidobacterium saeculare
Actinomyces cardiffensis
unclassified
Sample label (metadata):
c
l


2.7.8.6: Undecaprenyl-phosphate galactose phosphotransferase
Abundance (cpm)
50
40
30
20
10
0
60 samples
Contributions (linear scaling within total bar height):
Bifidobacterium
Bacteroides
Dialister
Roseburia
Anaerostipes
Catenibacterium
Eggerthella
Desulfovibrio
Coprococcus
Faecalibacterium
Acidaminococcus
Proteobacteria unclassified
Parasutterella
Hungatella
Veillonella
Mitsuokella
Intestinimonas
Streptococcus
other
unclassified
Sample label (metadata):
c
l


2.7.8.6: Undecaprenyl-phosphate galactose phosphotransferase
Abundance (cpm)
50
40
30
20
10
0
60 samples
Contributions (linear scaling within total bar height):
Bifidobacterium adolescentis
Bacteroides uniformis CAG 3
Bacteroides uniformis
Roseburia intestinalis
Dialister invisus
Anaerostipes hadrus
Dialister invisus CAG 218
Catenibacterium mitsuokai
Eggerthella lenta
Desulfovibrio piger
Coprococcus catus
Bifidobacterium longum
Faecalibacterium prausnitzii
Acidaminococcus fermentans
Bacteroides thetaiotaomicron
Proteobacteria bacterium CAG 139
Parasutterella excrementihominis
Bifidobacterium dentium
other
unclassified
Sample label (metadata):
c
l

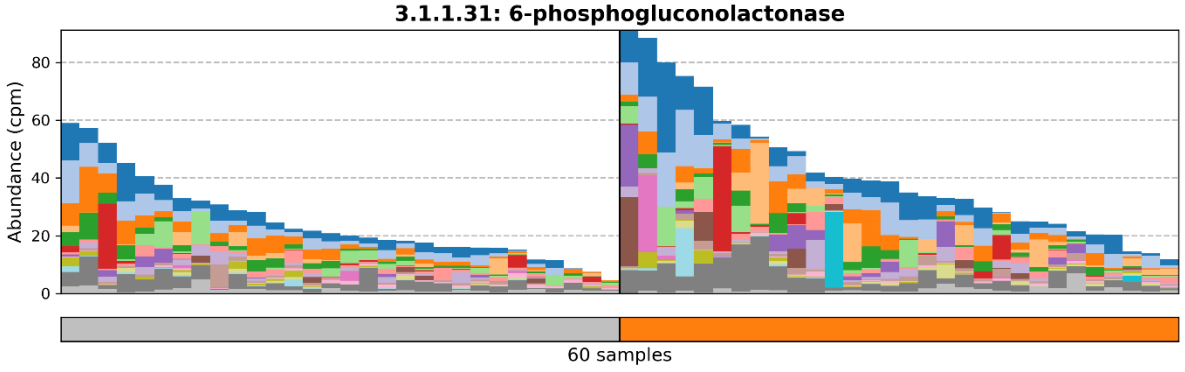



**Contributions (linear scaling within total bar height):**

- *Bacteroides*
- *Bifidobacterium*
- *Escherichia*
- *Parabacteroides*
- *Blautia*
- *Catenibacterium*
- *Roseburia*
- *Lactobacillus*
- *Coprococcus*
- *Desulfovibrio*
- *Eisenbergiella*
- *Veillonella*
- *Klebsiella*
- *Streptococcus*
- *Lachnoclostridium*
- *Haemophilus*
- *Hungatella*
- *Pediococcus*
- other
- unclassified

**Sample label (metadata):**

- c
- l

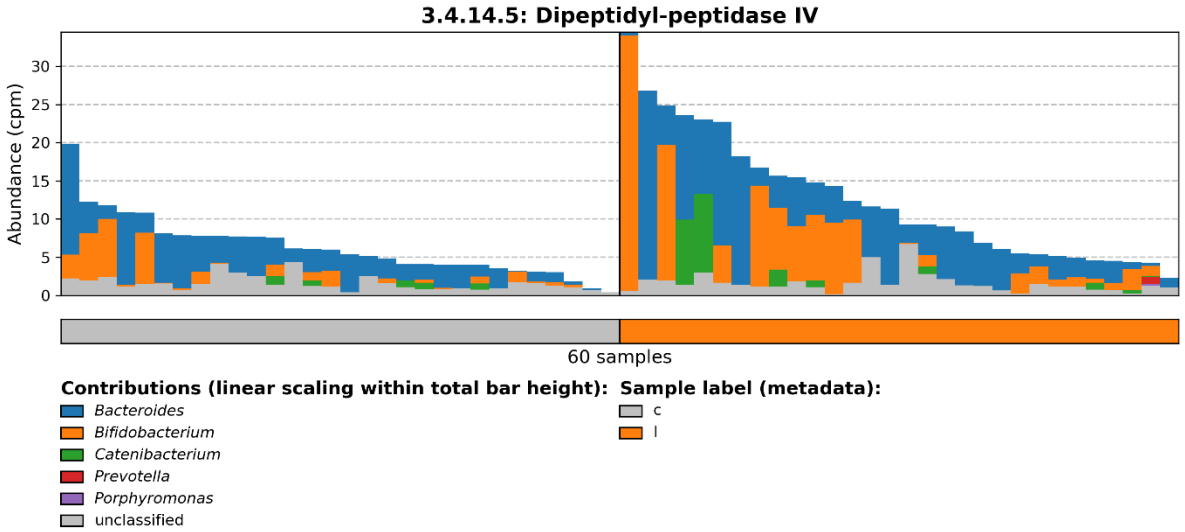


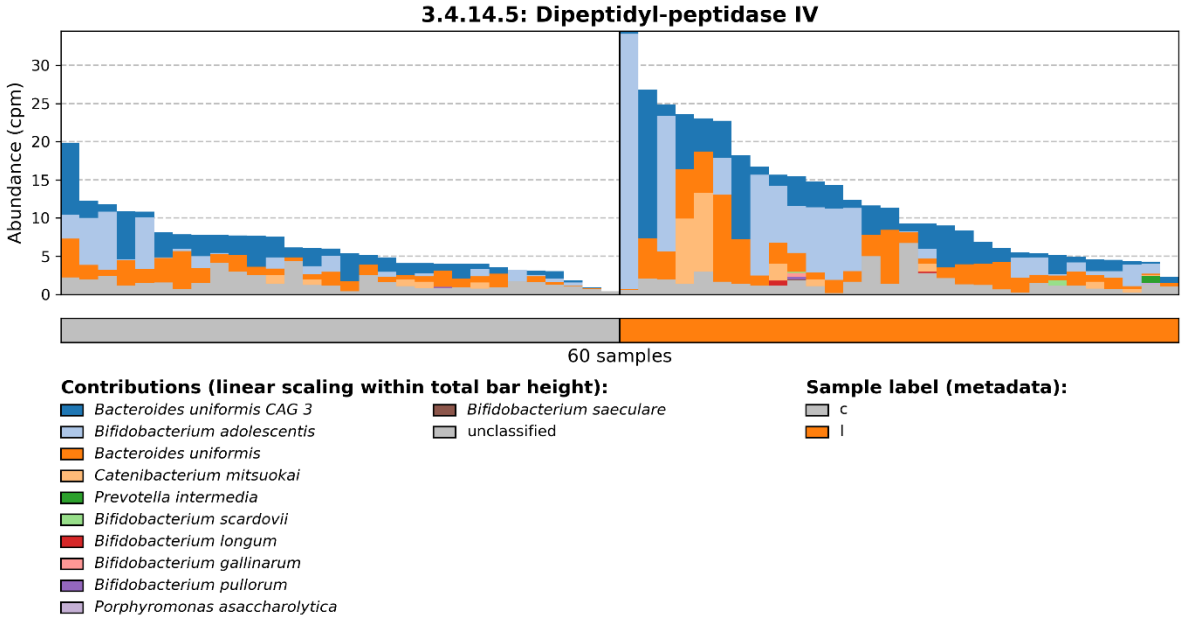

**Contributions (linear scaling within total bar height):**

- *Bacteroides uniformis*
- *Bacteroides uniformis CAG 3*
- *Bacteroides vulgatus CAG 6*
- *Bifidobacterium adolescentis*
- *Bacteroides vulgatus*
- *Bacteroides dorei*
- *Escherichia coli*
- *Blautia obeum*
- *Bacteroides cellulosilyticus*
- *Bacteroides thetaiotaomicron*
- *Catenibacterium mitsuokai*
- *Parabacteroides distasonis*
- *Bacteroides intestinalis*
- *Coprococcus catus*
- *Roseburia intestinalis*
- *Parabacteroides merdae*
- *Lactobacillus reuteri*
- *Bacteroides ovatus*
- other
- unclassified

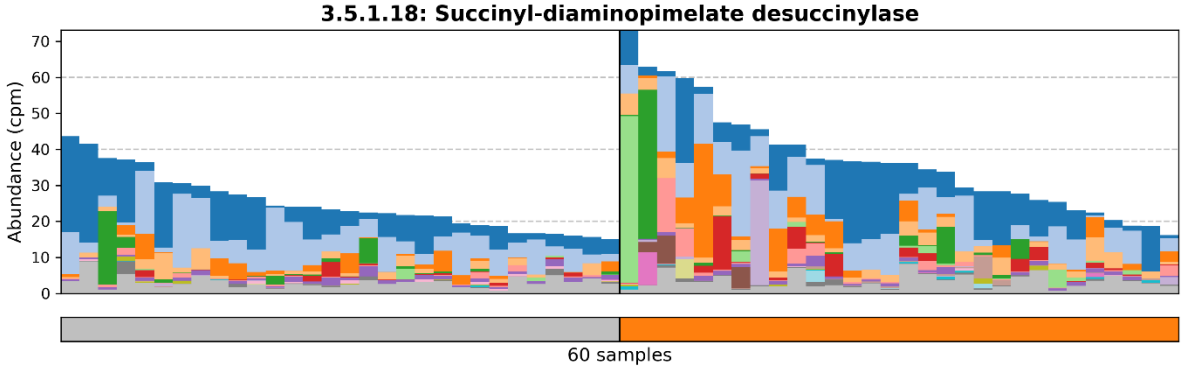
**Sample label (metadata):**

- c
- l


3.4.14.5: Dipeptidyl-peptidase IV
Abundance (cpm)
0
5
10
15
20
25
30
60 samples
Contributions (linear scaling within total bar height):
Bacteroides
Bifidobacterium
Catenibacterium
Prevotella
Porphyromonas
unclassified
Sample label (metadata):
c
l


3.4.14.5: Dipeptidyl-peptidase IV
Abundance (cpm)
0
5
10
15
20
25
30
60 samples
Contributions (linear scaling within total bar height):
Bacteroides uniformis CAG 3
Bifidobacterium adolescentis
Bacteroides uniformis
Catenibacterium mitsuokai
Prevotella intermedia
Bifidobacterium scardovii
Bifidobacterium longum
Bifidobacterium gallinarum
Bifidobacterium pullorum
Porphyromonas asaccharolytica
Bifidobacterium saeculare
unclassified
Sample label (metadata):
c
l

### 3.5.1.18: Succinyl-diaminopimelate desuccinylase



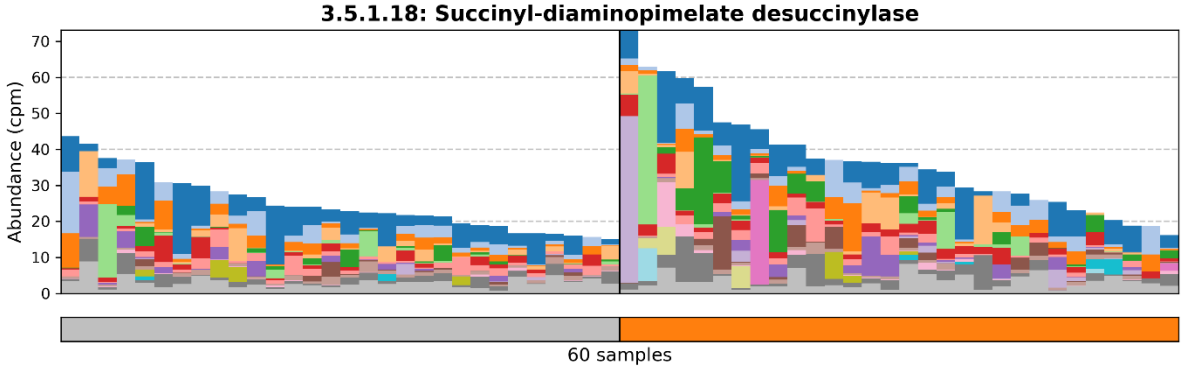

**Contributions (linear scaling within total bar height):**

- *Bacteroides*
- *Blautia*
- *Bifidobacterium*
- *Flavonifractor*
- *Escherichia*
- *Faecalitalea*
- *Ruminococcaceae unclassified*
- *Streptococcus*
- *Coprococcus*
- *Lactobacillus*
- *Pseudoflavonifractor*
- *Lachnoclostridium*
- *Klebsiella*
- *Parasutterella*
- *Proteobacteria unclassified*
- *Pediococcus*
- *Eisenbergiella*
- *Hungatella*
- other
- unclassified

**Sample label (metadata):**

- c
- l

### 3.5.1.18: Succinyl-diaminopimelate desuccinylase

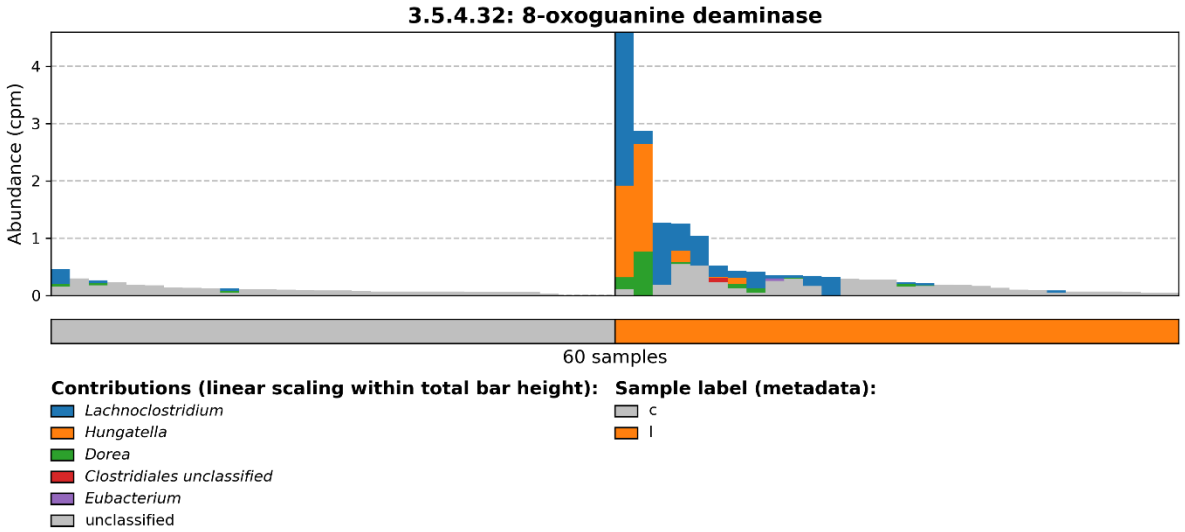


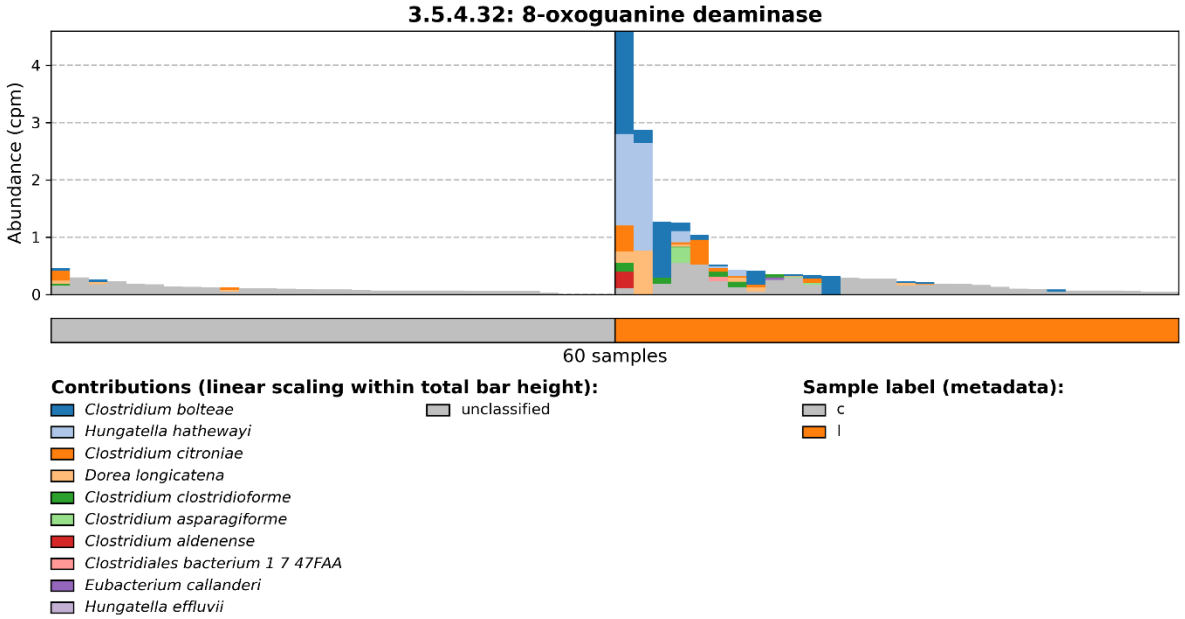

**Contributions (linear scaling within total bar height):**

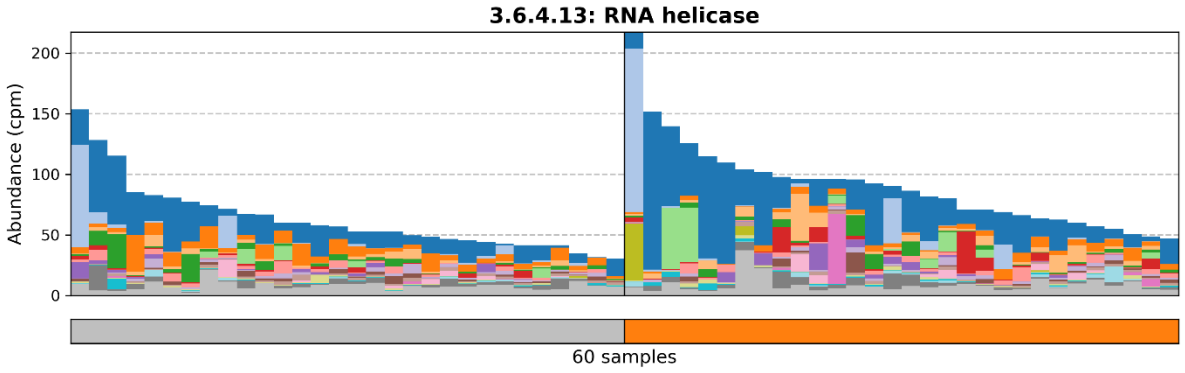
- *Ruminococcus torques*
- *Bacteroides vulgatus CAG 6*
- *Bacteroides vulgatus*
- *Bacteroides dorei*
- *Bifidobacterium adolescentis*
- *Escherichia coli*
- *Flavonifractor plautii*
- *Blautia obeum*
- *Bacteroides massiliensis*
- *Faecalitalea cylindroides*
- *Eubacterium siraeum*
- *Coprococcus catus*
- *Lactobacillus reuteri*
- *Streptococcus salivarius*
- *Bifidobacterium pseudocatenulatum*
- *Pseudoflavonifractor capillosus*
- *Flavonifractor sp An10*
- *Klebsiella pneumoniae*
- other
- unclassified

**Sample label (metadata):**

- c
- l


3.5.4.32: 8-oxoguanine deaminase
Abundance (cpm)
0
1
2
3
4
60 samples
Contributions (linear scaling within total bar height):
Lachnoclostridium
Hungatella
Dorea
Clostridiales unclassified
Eubacterium
unclassified
Sample label (metadata):
c
l


3.5.4.32: 8-oxoguanine deaminase
Abundance (cpm)
0
1
2
3
4
60 samples
Contributions (linear scaling within total bar height):
Clostridium bolteae
Hungatella hathewayi
Clostridium citroniae
Dorea longicatena
Clostridium clostridioforme
Clostridium asparagiforme
Clostridium aldenense
Clostridiales bacterium 1 7 47FAA
Eubacterium callanderi
Hungatella effluvii
unclassified
Sample label (metadata):
c
l

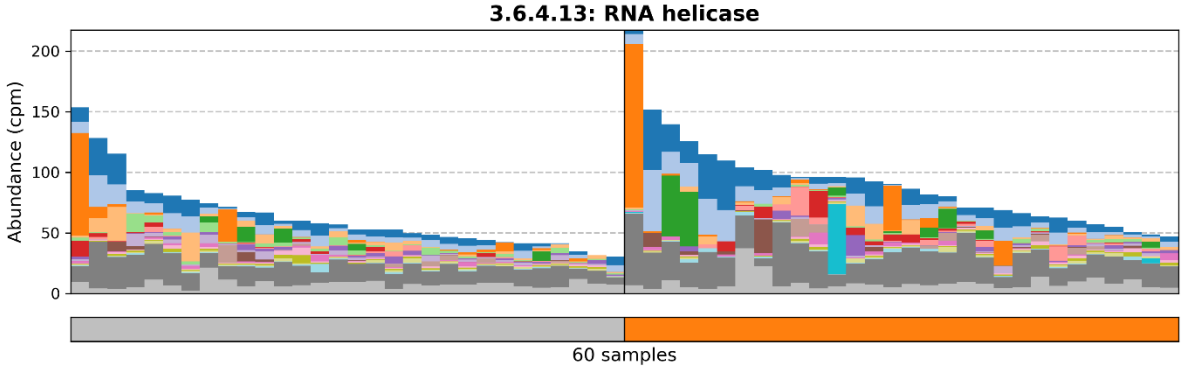



**Contributions (linear scaling within total bar height):**

- *Bacteroides*
- *Escherichia*
- *Eubacterium*
- *Bifidobacterium*
- *Lachnospiraceae unclassified*
- *Catenibacterium*
- *Streptococcus*
- *Roseburia*
- *Alistipes*
- *Dorea*
- *Blautia*
- *Coprococcus*
- *Lactobacillus*
- *Romboutsia*
- *Klebsiella*
- *Parabacteroides*
- *Flavonifractor*
- *Clostridium*
- other
- unclassified

**Sample label (metadata):**

- c
- l

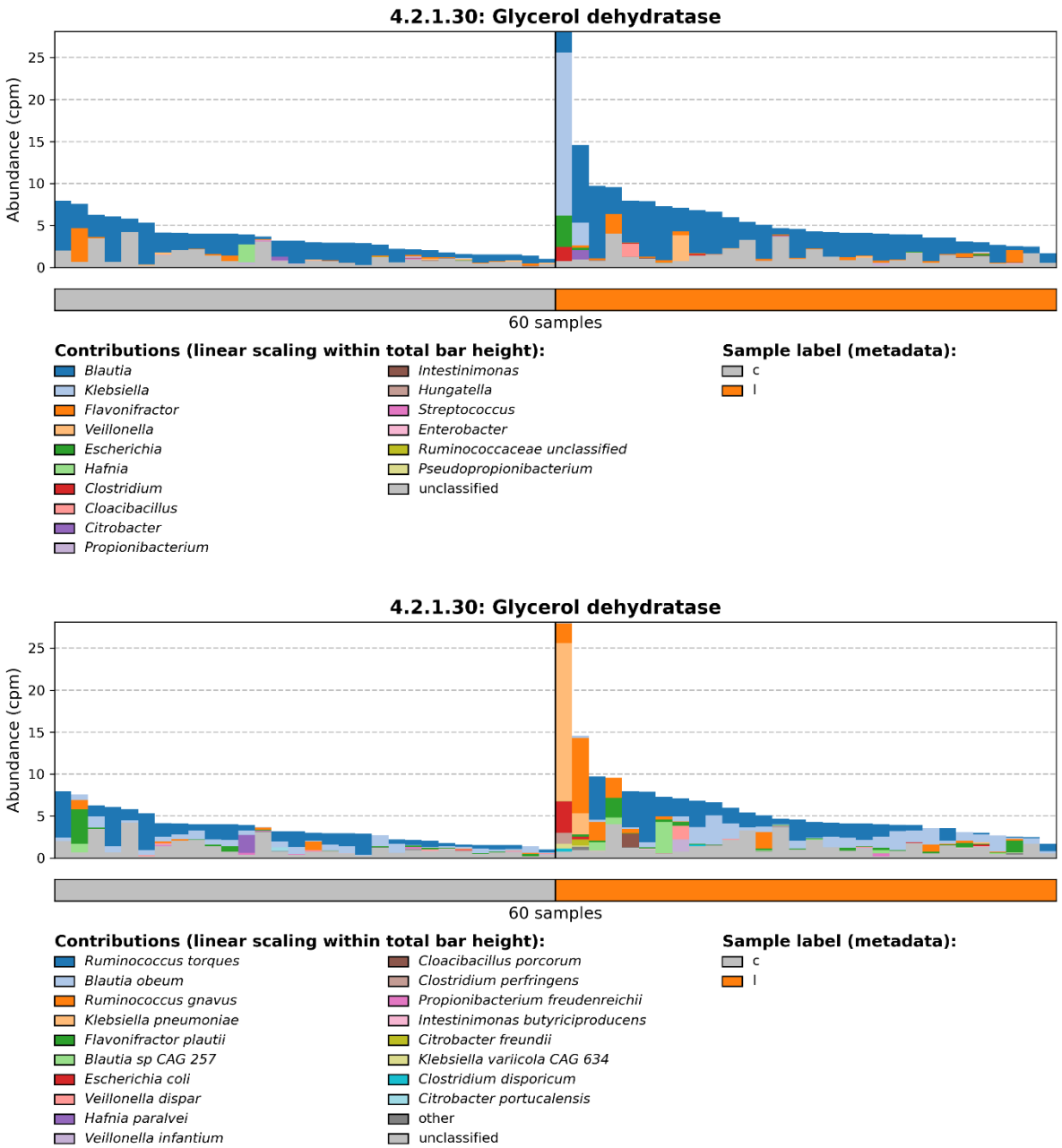



**Contributions (linear scaling within total bar height):**

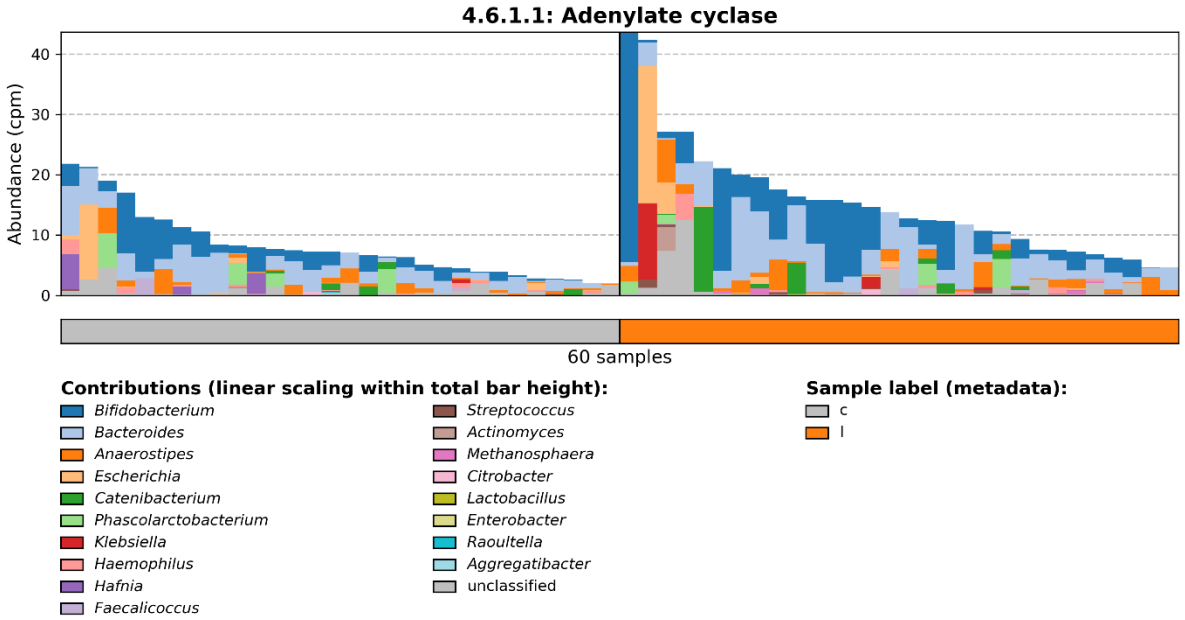
- *Bacteroides uniformis*
- *Bacteroides uniformis CAG 3*
- *Escherichia coli*
- *Eubacterium rectale*
- *Catenibacterium mitsuokai*
- *Eubacterium eligens CAG 72*
- *Alistipes finegoldii*
- *Bifidobacterium adolescentis*
- *Ruminococcus torques*
- *Eubacterium eligens*
- *Bacteroides eggerthii*
- *Romboutsia ilealis*
- *Bacteroides thetaiotaomicron*
- *Eubacterium hallii*
- *Eubacterium hallii CAG 12*
- *Dorea longicatena CAG 42*
- *Lactobacillus reuteri*
- *Parabacteroides distasonis*
- other
- unclassified

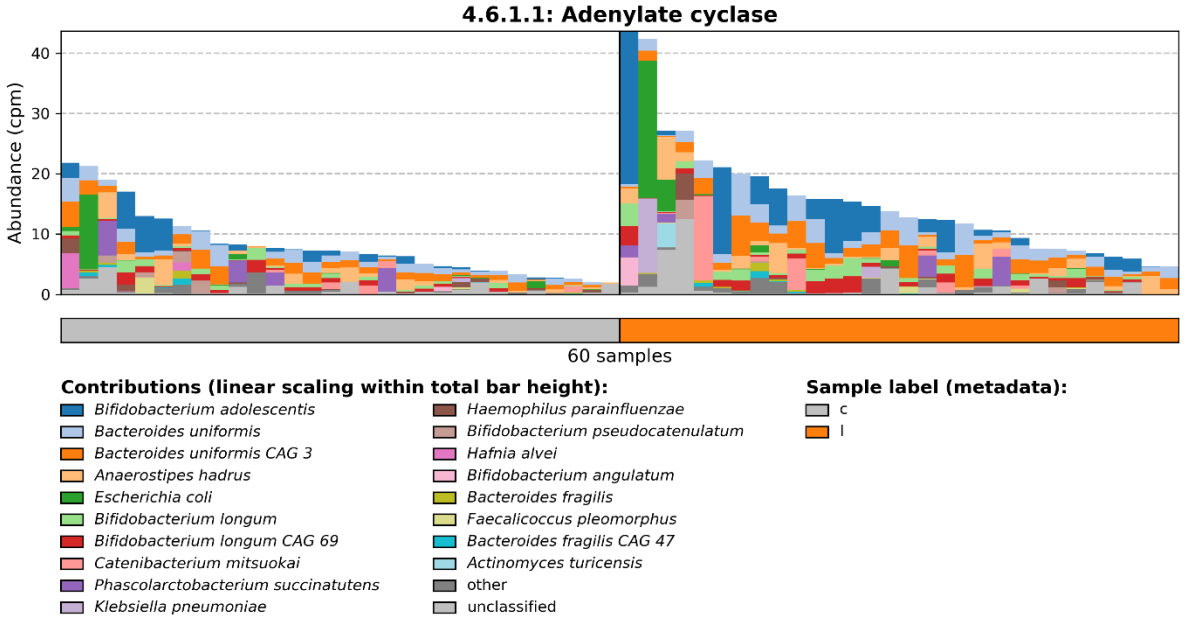
**Sample label (metadata):**

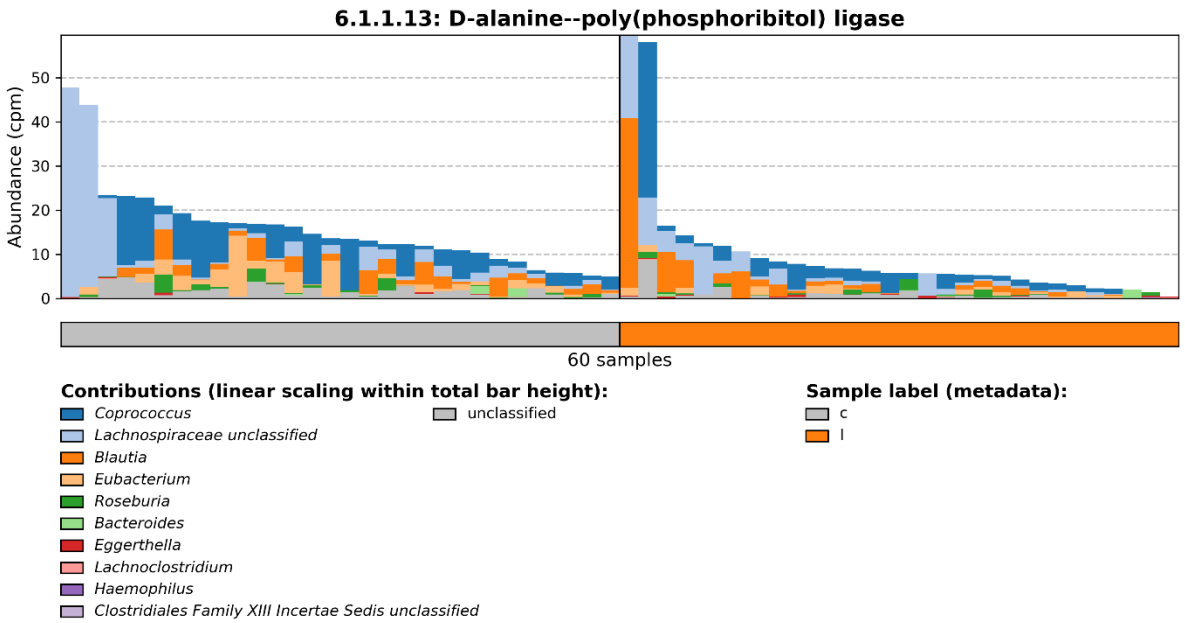
- c
- l

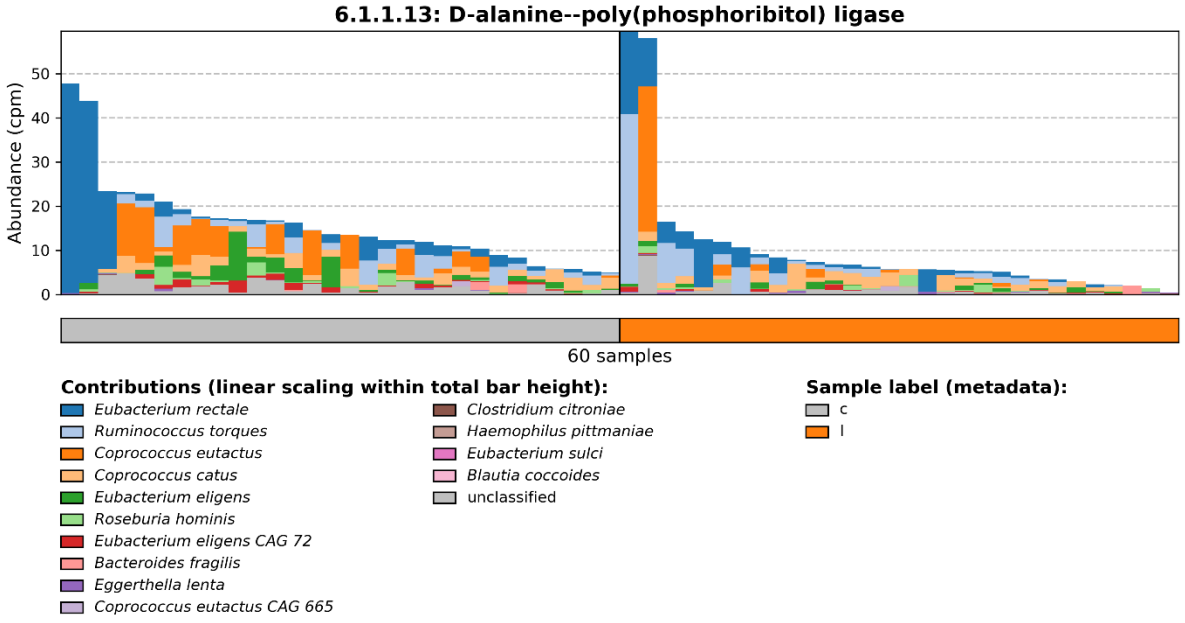

4.2.1.30: Glycerol dehydratase
Abundance (cpm)
25
20
15
10
5
0
60 samples
Contributions (linear scaling within total bar height):
Blautia
Klebsiella
Flavonifractor
Veillonella
Escherichia
Hafnia
Clostridium
Cloacibacillus
Citrobacter
Propionibacterium
Intestinimonas
Hungatella
Streptococcus
Enterobacter
Ruminococcaceae unclassified
Pseudopropionibacterium
unclassified
Sample label (metadata):
c
l

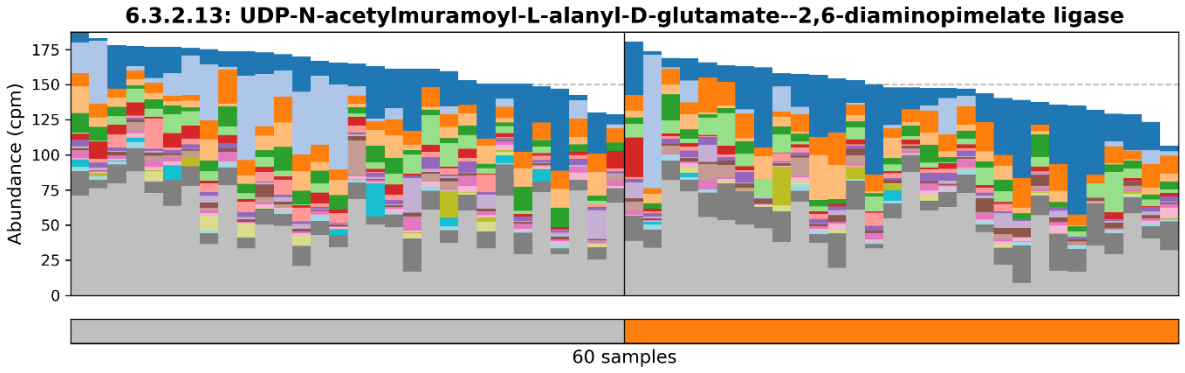

4.2.1.30: Glycerol dehydratase
Abundance (cpm)
25
20
15
10
5
0
60 samples
Contributions (linear scaling within total bar height):
Ruminococcus torques
Blautia obeum
Ruminococcus gnavus
Klebsiella pneumoniae
Flavonifractor plautii
Blautia sp CAG 257
Escherichia coli
Veillonella dispar
Hafnia paralvei
Veillonella infantium
Cloacibacillus porcorum
Clostridium perfringens
Propionibacterium freudenreichii
Intestinimonas butyriciproducens
Citrobacter freundii
Klebsiella variicola CAG 634
Clostridium disporicum
Citrobacter portucalensis
other
unclassified
Sample label (metadata):
c
l


4.6.1.1: Adenylate cyclase
Abundance (cpm)
0
10
20
30
40
60 samples
Contributions (linear scaling within total bar height):
Bifidobacterium
Bacteroides
Anaerostipes
Escherichia
Catenibacterium
Phascolarctobacterium
Klebsiella
Haemophilus
Hafnia
Faecalicoccus
Streptococcus
Actinomyces
Methanosphaera
Citrobacter
Lactobacillus
Enterobacter
Raoultella
Aggregatibacter
unclassified
Sample label (metadata):
c
l


4.6.1.1: Adenylate cyclase
Abundance (cpm)
0
10
20
30
40
60 samples
Contributions (linear scaling within total bar height):
Bifidobacterium adolescentis
Bacteroides uniformis
Bacteroides uniformis CAG 3
Anaerostipes hadrus
Escherichia coli
Bifidobacterium longum
Bifidobacterium longum CAG 69
Catenibacterium mitsuokai
Phascolarctobacterium succinatutens
Klebsiella pneumoniae
Haemophilus parainfluenzae
Bifidobacterium pseudocatenulatum
Hafnia alvei
Bifidobacterium angulatum
Bacteroides fragilis
Faecalicoccus pleomorphus
Bacteroides fragilis CAG 47
Actinomyces turicensis
other
unclassified
Sample label (metadata):
c
l


6.1.1.13: D-alanine--poly(phosphoribitol) ligase
Abundance (cpm)
0
10
20
30
40
50
60 samples
Contributions (linear scaling within total bar height):
Coprococcus
Lachnospiraceae unclassified
Blautia
Eubacterium
Roseburia
Bacteroides
Eggerthella
Lachnoclostridium
Haemophilus
Clostridiales Family XIII Incertae Sedis unclassified
unclassified
Sample label (metadata):
c
l


6.1.1.13: D-alanine--poly(phosphoribitol) ligase
Abundance (cpm)
0
10
20
30
40
50
60 samples
Contributions (linear scaling within total bar height):
Eubacterium rectale
Ruminococcus torques
Coprococcus eutactus
Coprococcus catus
Eubacterium eligens
Roseburia hominis
Eubacterium eligens CAG 72
Bacteroides fragilis
Eggerthella lenta
Coprococcus eutactus CAG 665
Clostridium citroniae
Haemophilus pittmaniae
Eubacterium sulci
Blautia coccoides
unclassified
Sample label (metadata):
c
l

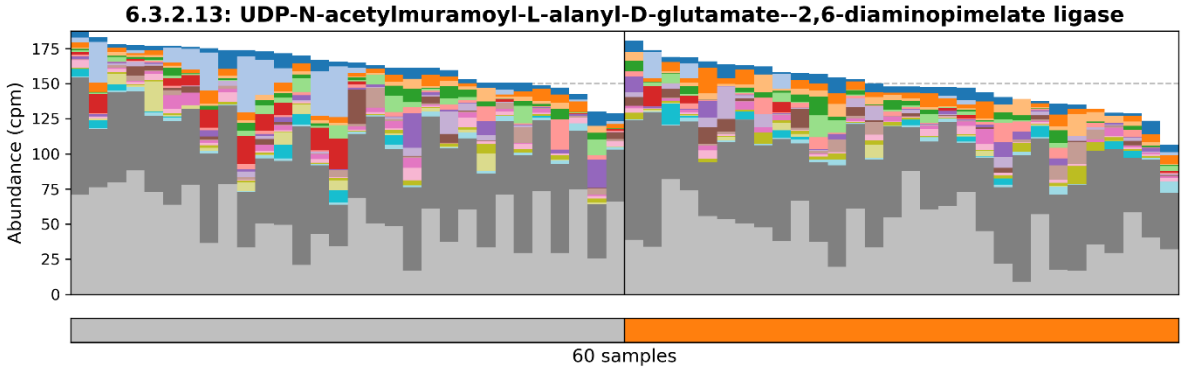



**Contributions (linear scaling within total bar height):**

- *Bacteroides*
- *Prevotella*
- *Blautia*
- *Roseburia*
- *Eubacterium*
- *Alistipes*
- *Coprococcus*
- *Ruminococcus*
- *Dorea*
- *Lachnospiraceae unclassified*
- *Parabacteroides*
- *Akkermansia*
- *Anaerostipes*
- *Fusicatenibacter*
- *Escherichia*
- *Dialister*
- *Butyrivibrio*
- *Phascolarctobacterium*
- other
- unclassified

**Sample label (metadata):**

- c
- l



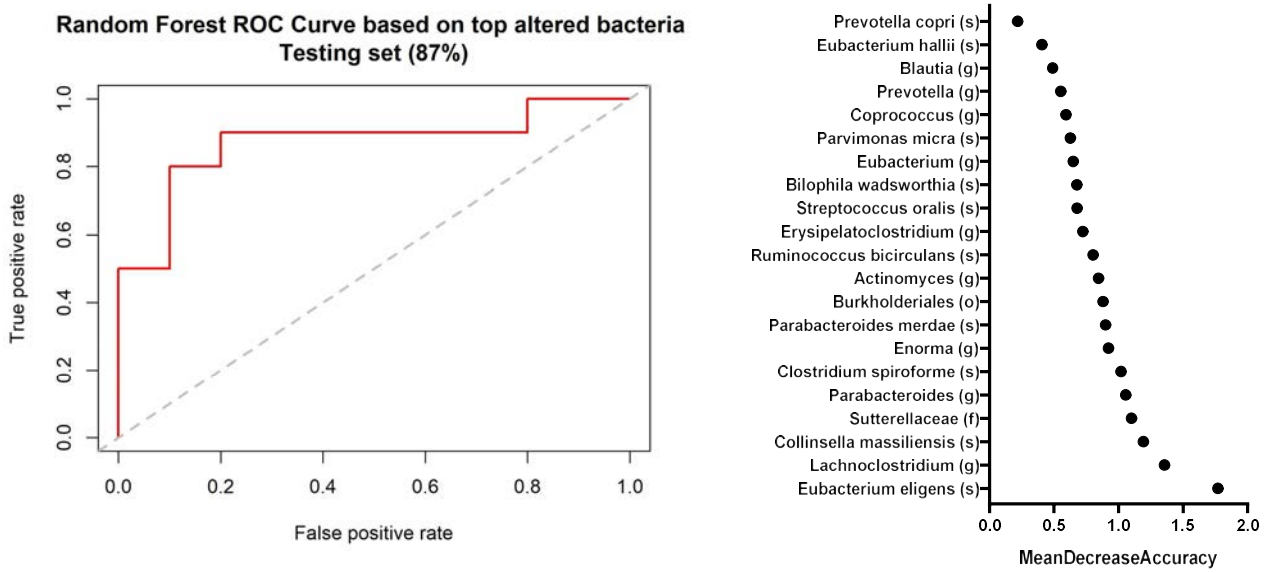

**Contributions (linear scaling within total bar height):**

- *Blautia obeum*
- *Prevotella copri CAG 164*
- *Ruminococcus torques*
- *Bacteroides uniformis*
- *Bacteroides vulgatus CAG 6*
- *Roseburia faecis*
- *Prevotella copri*
- *Bacteroides vulgatus*
- *Eubacterium rectale*
- *Alistipes finegoldii*
- *Akkermansia muciniphila*
- *Bacteroides dorei*
- *Alistipes putredinis*
- *Anaerostipes hadrus*
- *Bacteroides uniformis CAG 3*
- *Ruminococcus bicirculans*
- *Eubacterium sp CAG 180*
- *Fusicatenibacter saccharivorans*
- other
- unclassified

**Sample label (metadata):**

- c
- l

A




B

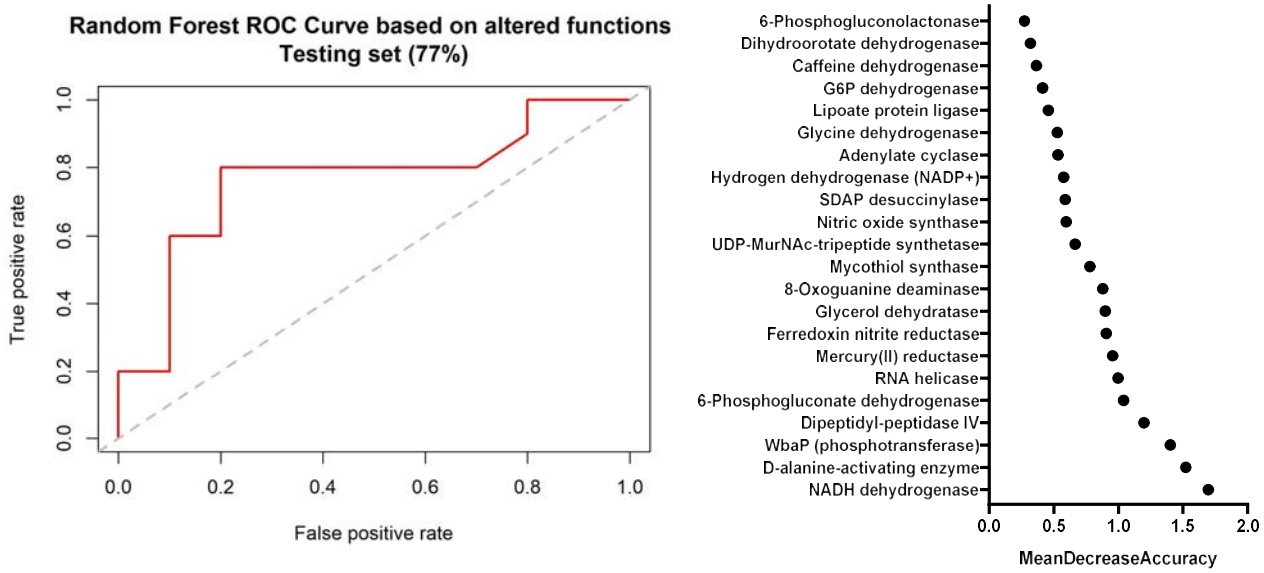



**Figure S5. Bacterial taxa and functional signatures identified in AML patients.**

A) RandomForest model ROC curve based on the top altered bacteria to predict AML status.

B) RandomForest model ROC curve based on altered EC enzyme functions to predict AML status.

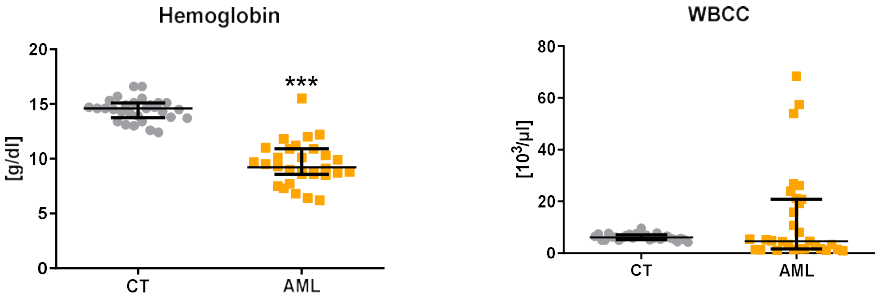



**Figure S6. Confirmation of hemoglobin and white blood cell count alteration in AML patients.**
Hemoglobin and WBCC (white blood cell count).
Variables are non-normally distributed and are expressed as median (interquartile range) and are tested by a Mann-Whitney U-test. AML in orange *vs.* CT in grey.
*** : p-value < 0.001

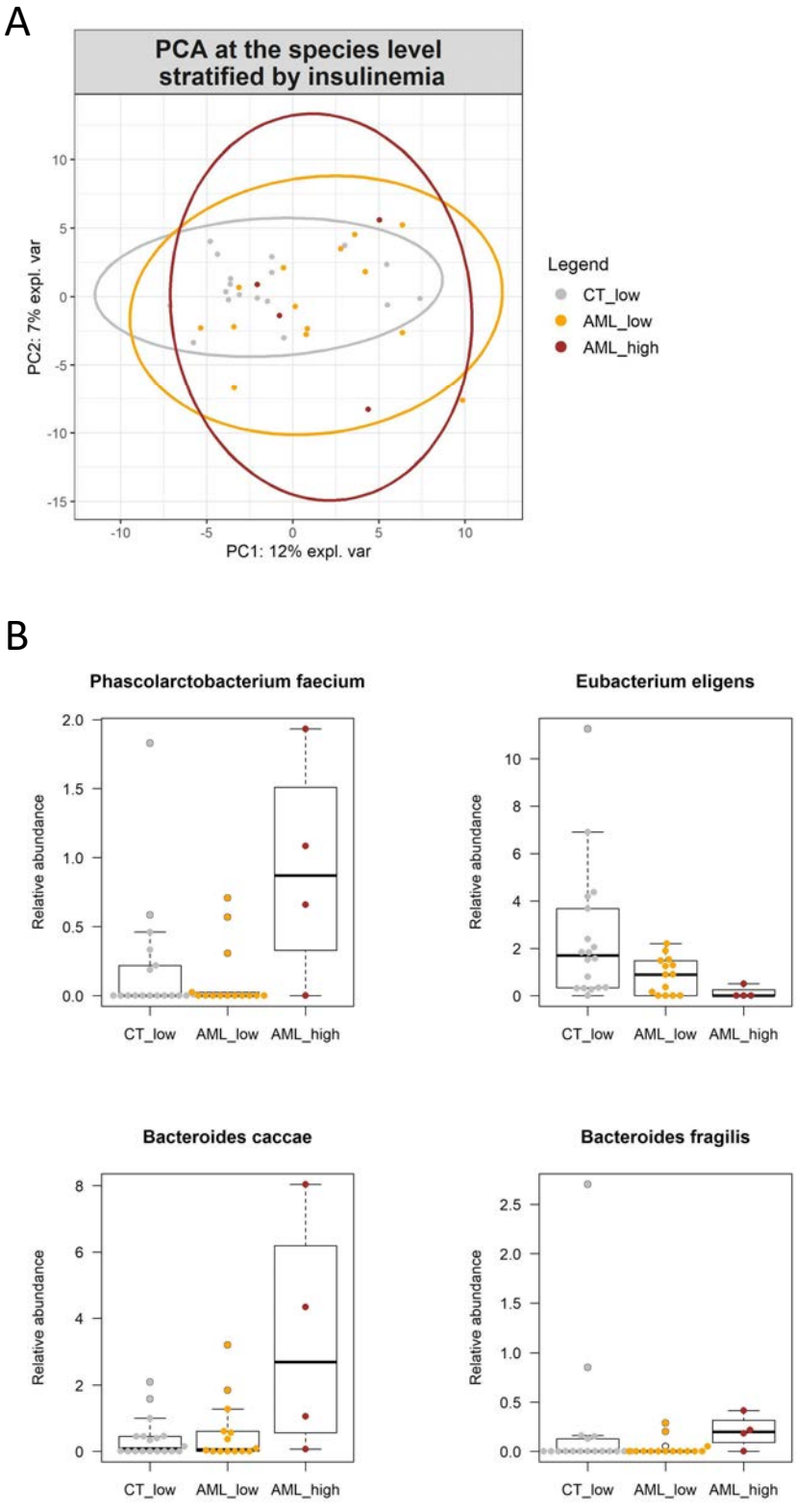



**Figure S7. Alterations in the gut microbiota composition in AML patients with high insulinemia (results of metagenomics sequencing).**

A) PCA analyses on CLR-transformed data at the species level, stratified by insulinemia and disease (CT low/AML low/ AML high). Insulinemia class does not explain a significant part of the variance in the dataset (PERMANOVA ns). B) 4 species were significantly different between individuals with low versus high insulinemia levels (pvalue<0.05, qvalue ns).

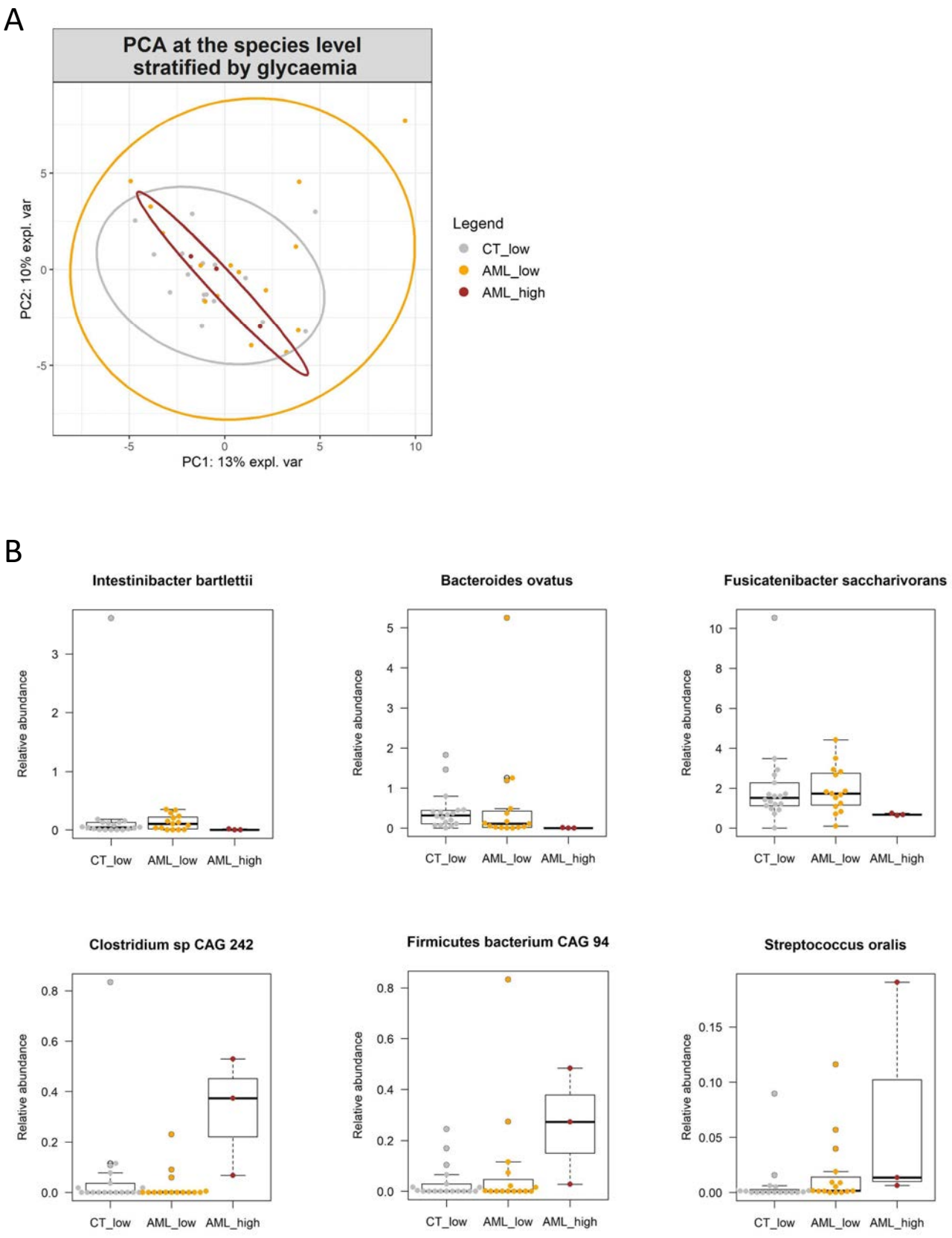



**Figure S8. Alterations in the gut microbiota composition in AML patients with high glycaemia (results of metagenomics sequencing).**

A) PCA analyses on CLR-transformed data at the species level, stratified by glycaemia and disease (CT low/AML low/ AML high). Glycaemia class does not explain a significant part of the variance in the dataset (PERMANOVA ns). B) 6 species were significantly different between individuals with low versus high glycemia levels (pvalue<0.05, qvalue ns).

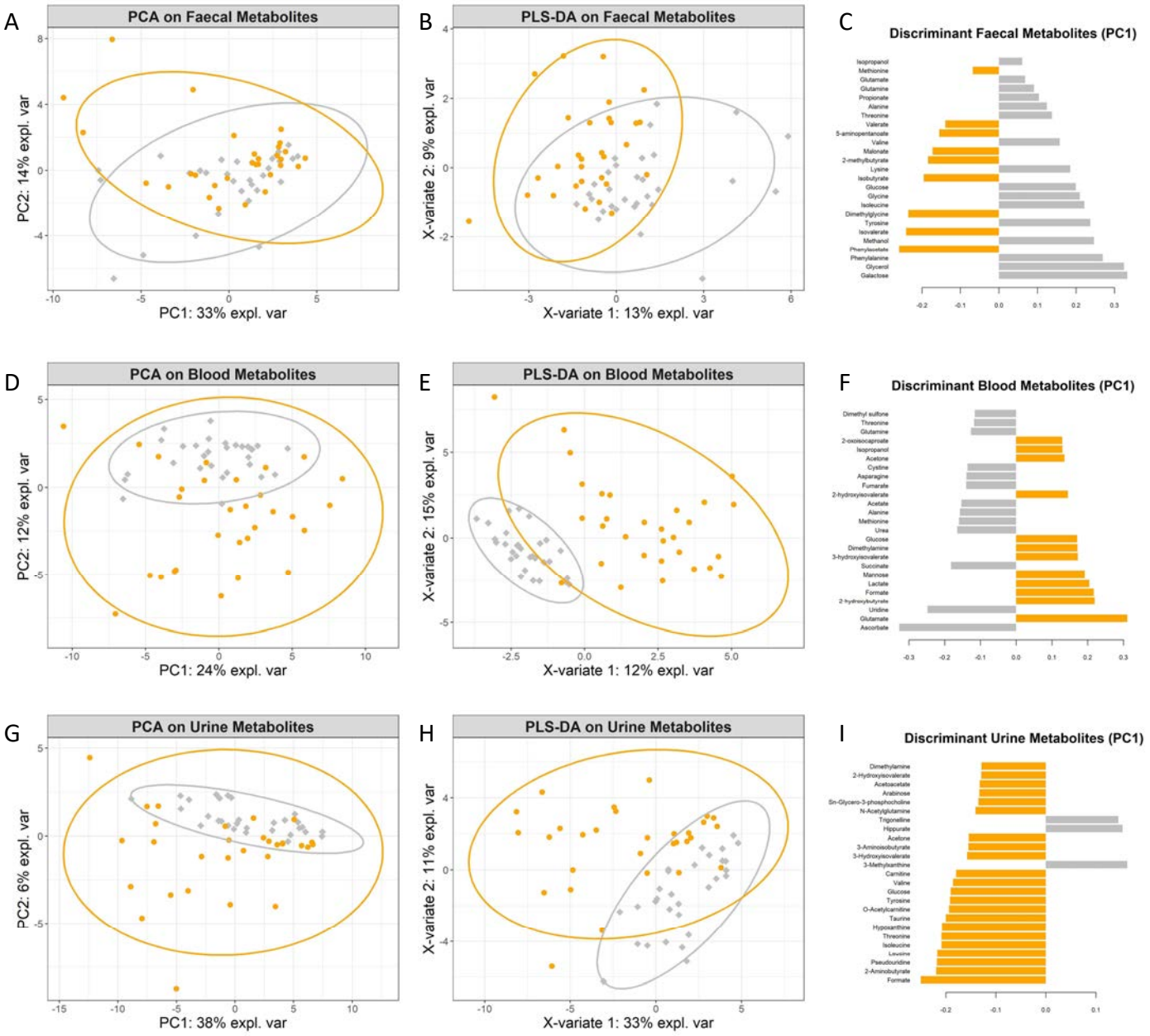



**Figure S9. Multivariate analyses on faecal, blood and urine metabolites pinpoint differences between AML and CT subjects.**

Principal component analyses (PCA), partial least square discriminant analyses (PLS-DA) and first 25 loadings of the PLS-DA for faecal (A-C), blood (D-F) and urine (G-I) metabolites (first principal component). A) PERMANOVA: R2 = 0.2% ns. D) PERMANOVA: R2 = 11.2% ** G) PERMANOVA R2 = 2.9% *. AML in orange vs. CT in grey. *p-value < 0.05; **p-value < 0.01.

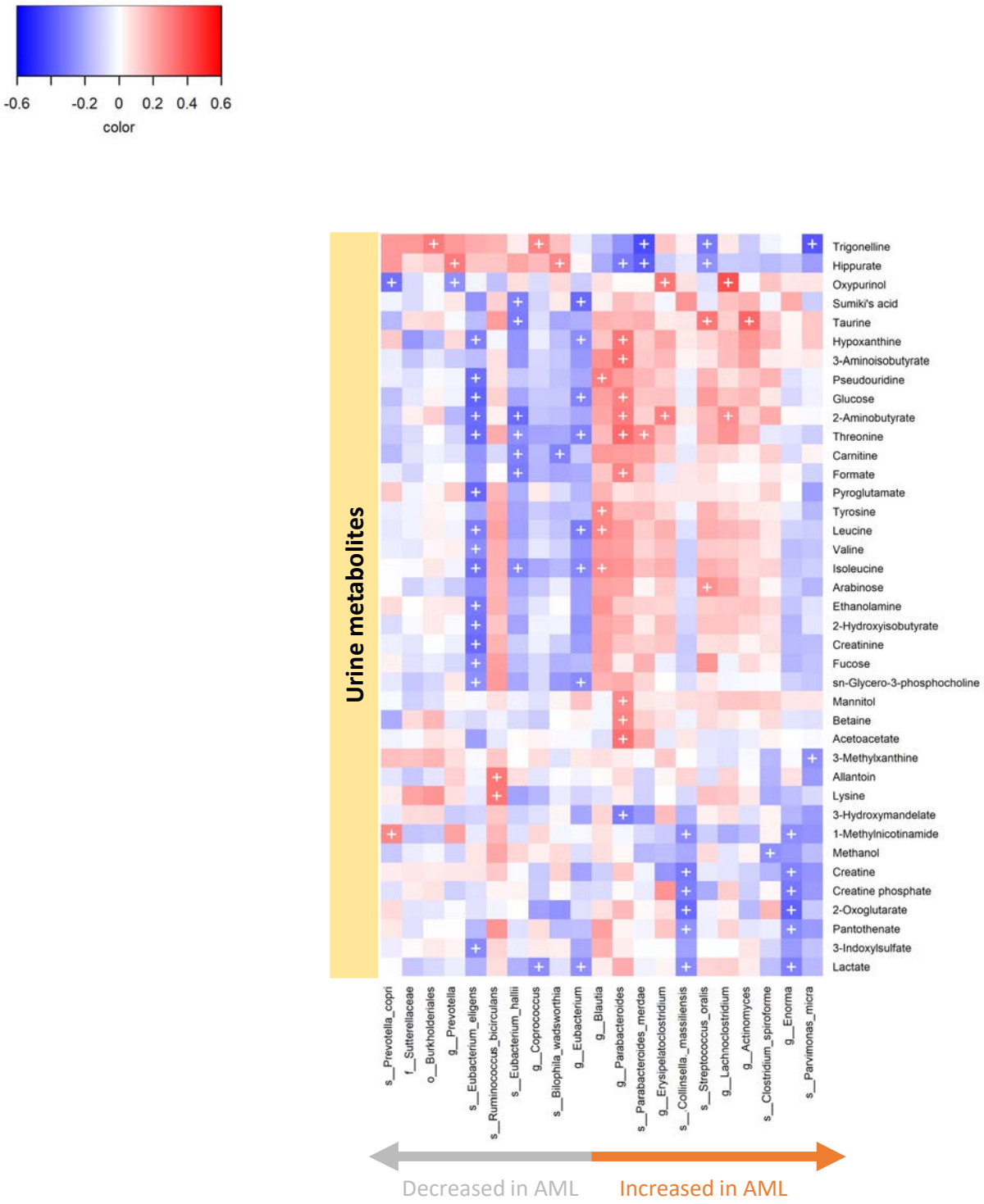
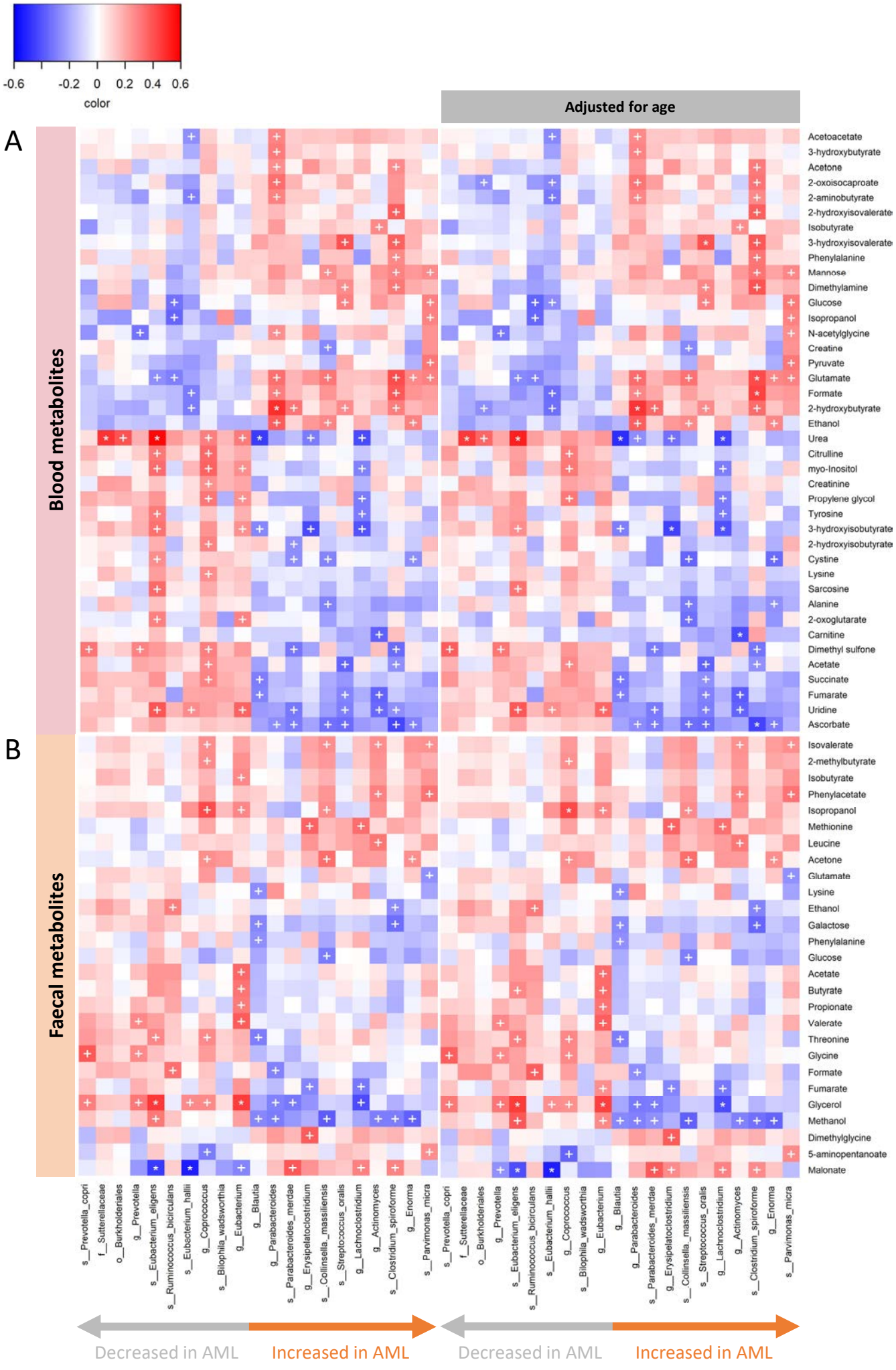



**Figure S10. Correlations between urine metabolites and the top altered bacteria.**
Spearman correlations. Metabolites with at least one correlation with an altered taxon are present. Microbial taxa are ordered by fold change. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1.


-0.6
-0.2
0
0.2
0.4
0.6
color
Adjusted for age
A
Blood metabolites
Acetoacetate
3-hydroxybutyrate
Acetone
2-oxoisocaproate
2-aminobutyrate
2-hydroxyisovalerate
Isobutyrate
3-hydroxyisovalerate
Phenylalanine
Mannose
Dimethylamine
Glucose
Isopropanol
N-acetylglycine
Creatine
Pyruvate
Glutamate
Formate
2-hydroxybutyrate
Ethanol
Urea
Citrulline
myo-Inositol
Creatinine
Propylene glycol
Tyrosine
3-hydroxyisobutyrate
2-hydroxyisobutyrate
Cystine
Lysine
Sarcosine
Alanine
2-oxoglutarate
Carnitine
Dimethyl sulfone
Acetate
Succinate
Fumarate
Uridine
Ascorbate
B
Faecal metabolites
Isovalerate
2-methylbutyrate
Isobutyrate
Phenylacetate
Isopropanol
Methionine
Leucine
Acetone
Glutamate
Lysine
Ethanol
Galactose
Phenylalanine
Glucose
Acetate
Butyrate
Propionate
Valerate
Threonine
Glycine
Formate
Fumarate
Glycerol
Methanol
Dimethylglycine
5-aminopentanoate
Malonate
s__Prevotella_copri
f__Sutterellaceae
o__Burkholderiales
g__Prevotella
s__Eubacterium_eligens
s__Ruminococcus_bicirculans
s__Eubacterium_hallii
g__Coprococcus
s__Bilophila_wadsworthia
g__Eubacterium
g__Blautia
g__Parabacteroides
s__Parabacteroides_merdae
g__Erysipelatoclostridium
s__Collinsella_massiliensis
s__Streptococcus_oralis
g__Lachnoclostridium
g__Actinomyces
s__Clostridium_spiroforme
g__Enorma
s__Parvimonas_micra
Decreased in AML
Increased in AML
Decreased in AML
Increased in AML

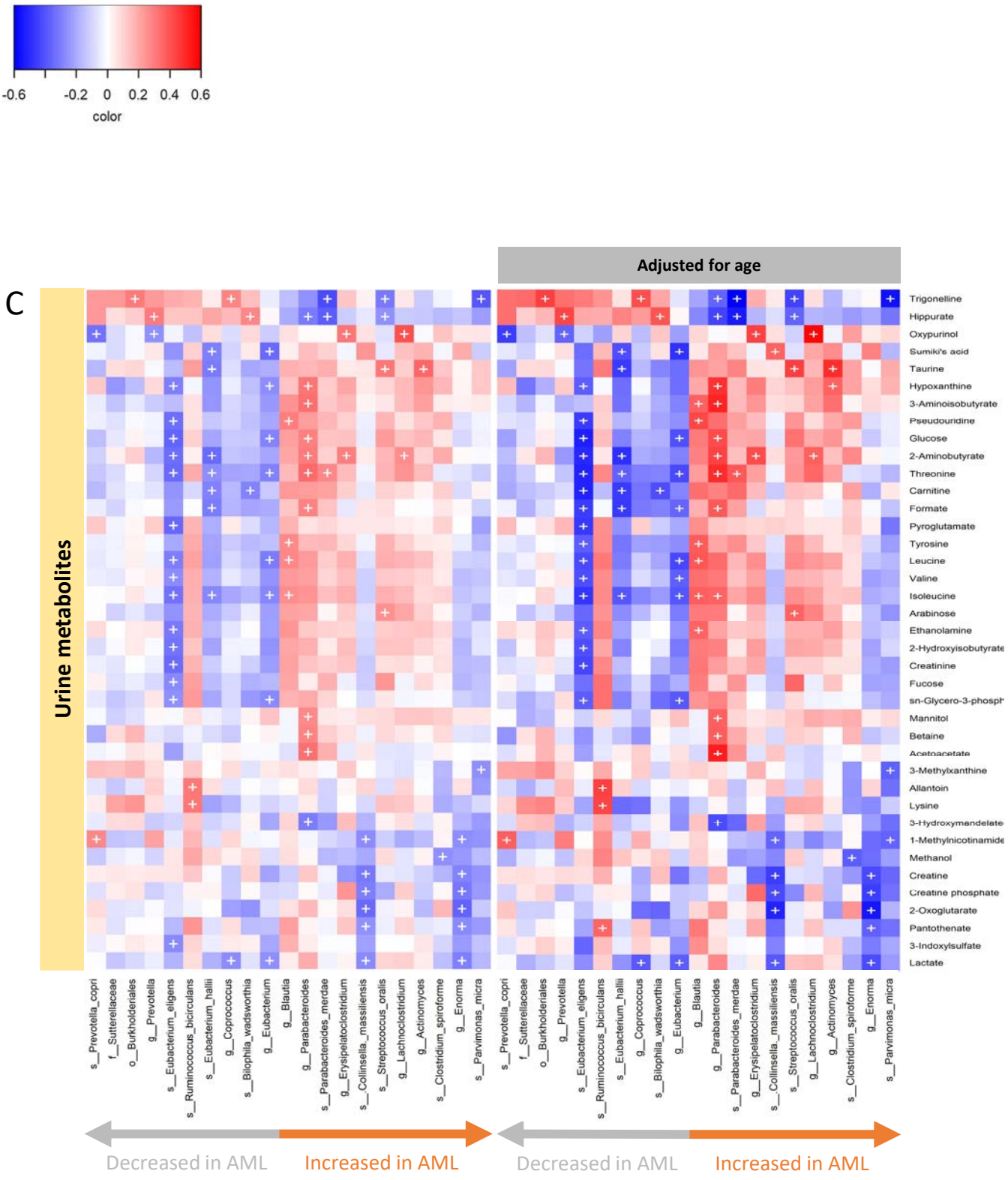



**Figure S11. Correlations between blood, faecal and urine metabolites and the top altered bacteria.** Spearman correlations (left) and partial Spearman rank-based correlations (pSRBC) adjusted for age (right) for the whole cohort (AML group and CT group). Metabolites with at least one correlation with an altered taxon are present. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1. A) Correlations between blood metabolites and top altered bacteria. B) Correlations between faecal metabolites and top altered bacteria. C) Correlations between urine metabolites and top altered bacteria.

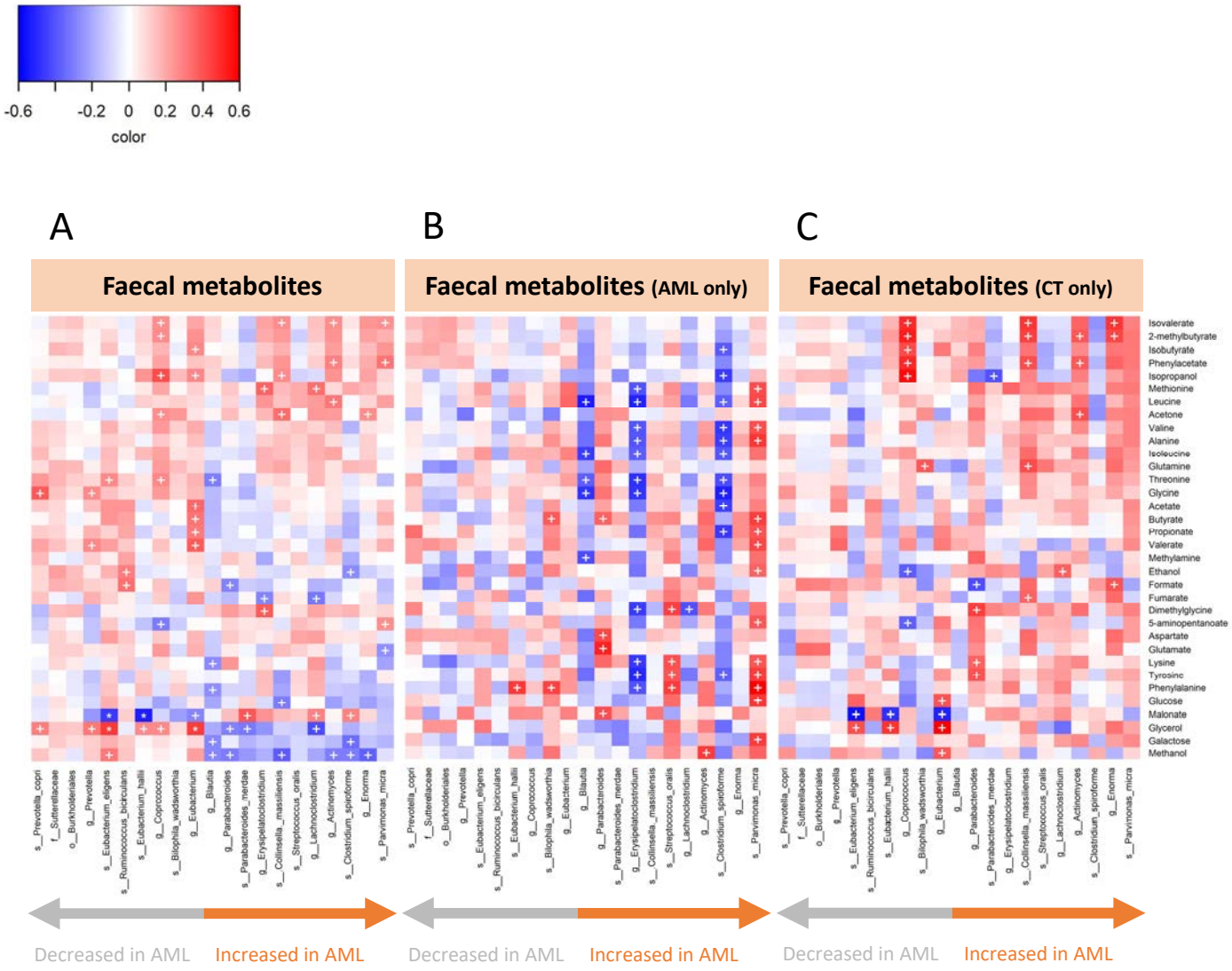



**Figure S12. Correlations between faecal metabolites and the top altered bacteria.**
Spearman correlations. All identified metabolites are present. Microbial taxa are ordered by fold change in the whole dataset. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1. A) Spearman correlations within the whole cohort, both acute myeloid leukaemia group (AML group) and the healthy control group (CT group). B) Spearman correlations within the acute myeloid leukaemia group (AML group). C) Spearman correlations within the healthy control group (CT group).

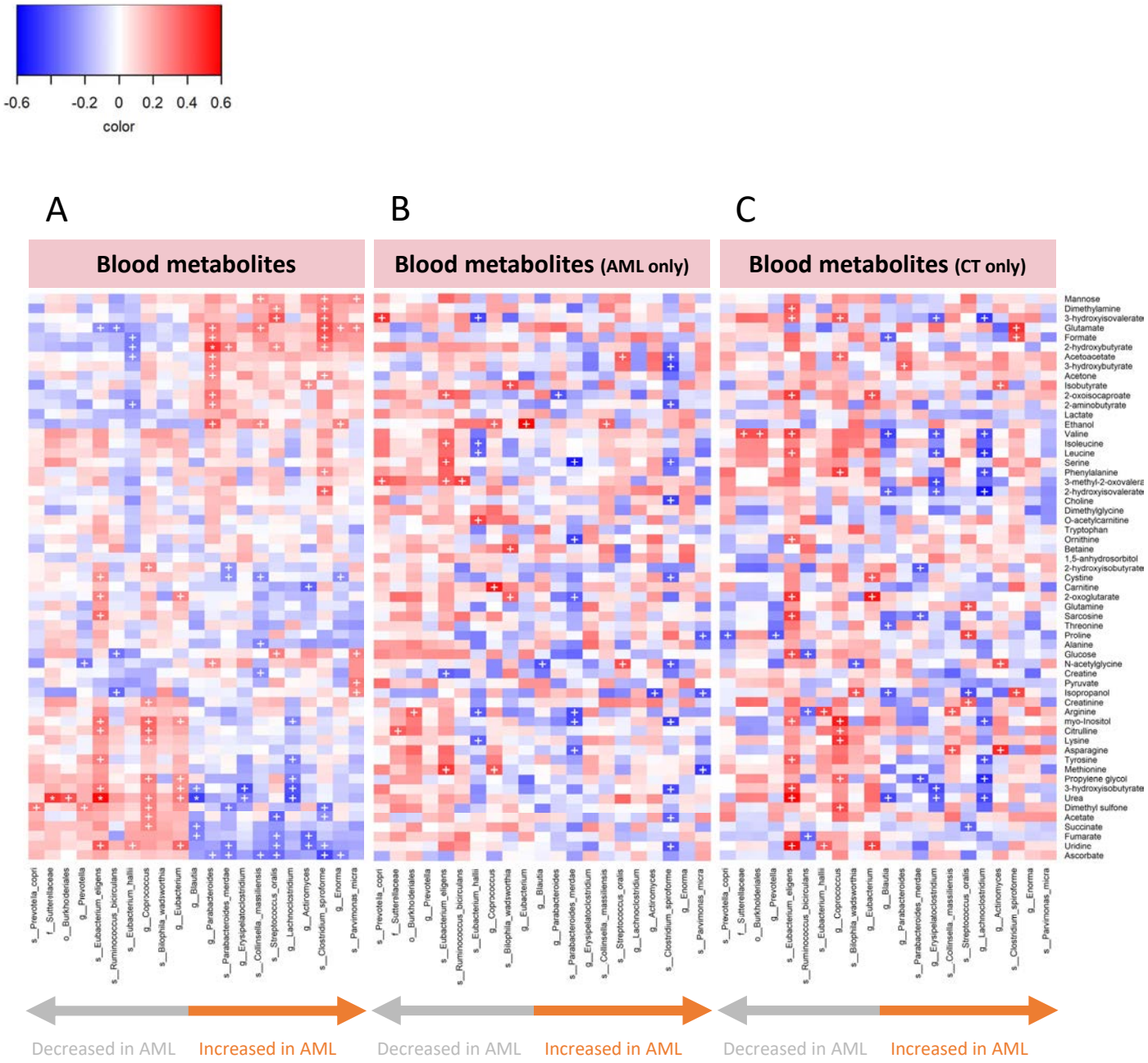



**Figure S13. Correlations between blood metabolites and the top altered bacteria.**
Spearman correlations. All identified metabolites are present. Microbial taxa are ordered by fold change in the whole dataset. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1. A) Spearman correlations within the whole cohort, both acute myeloid leukaemia group (AML group) and the healthy control group (CT group). B) Spearman correlations within the acute myeloid leukaemia group (AML group). C) Spearman correlations within the healthy control group (CT group).

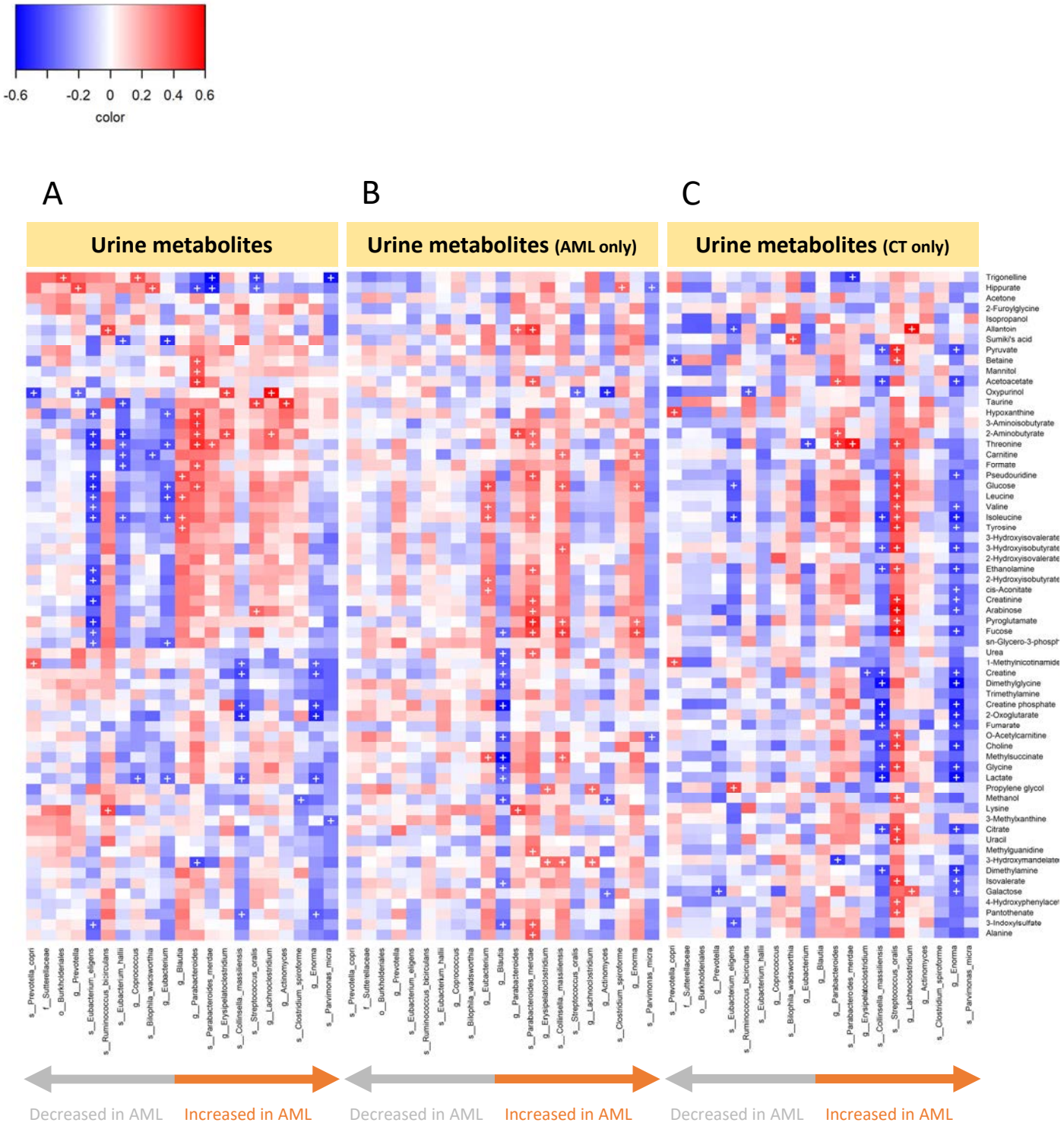



**Figure S14. Correlations between urine metabolites and the top altered bacteria.**
Spearman correlations. All identified metabolites are present. Microbial taxa are ordered by fold change in the whole dataset. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1. A) Spearman correlations within the whole cohort, both acute myeloid leukaemia group (AML group) and the healthy control group (CT group). B) Spearman correlations within the acute myeloid leukaemia group (AML group). C) Spearman correlations within the healthy control group (CT group).

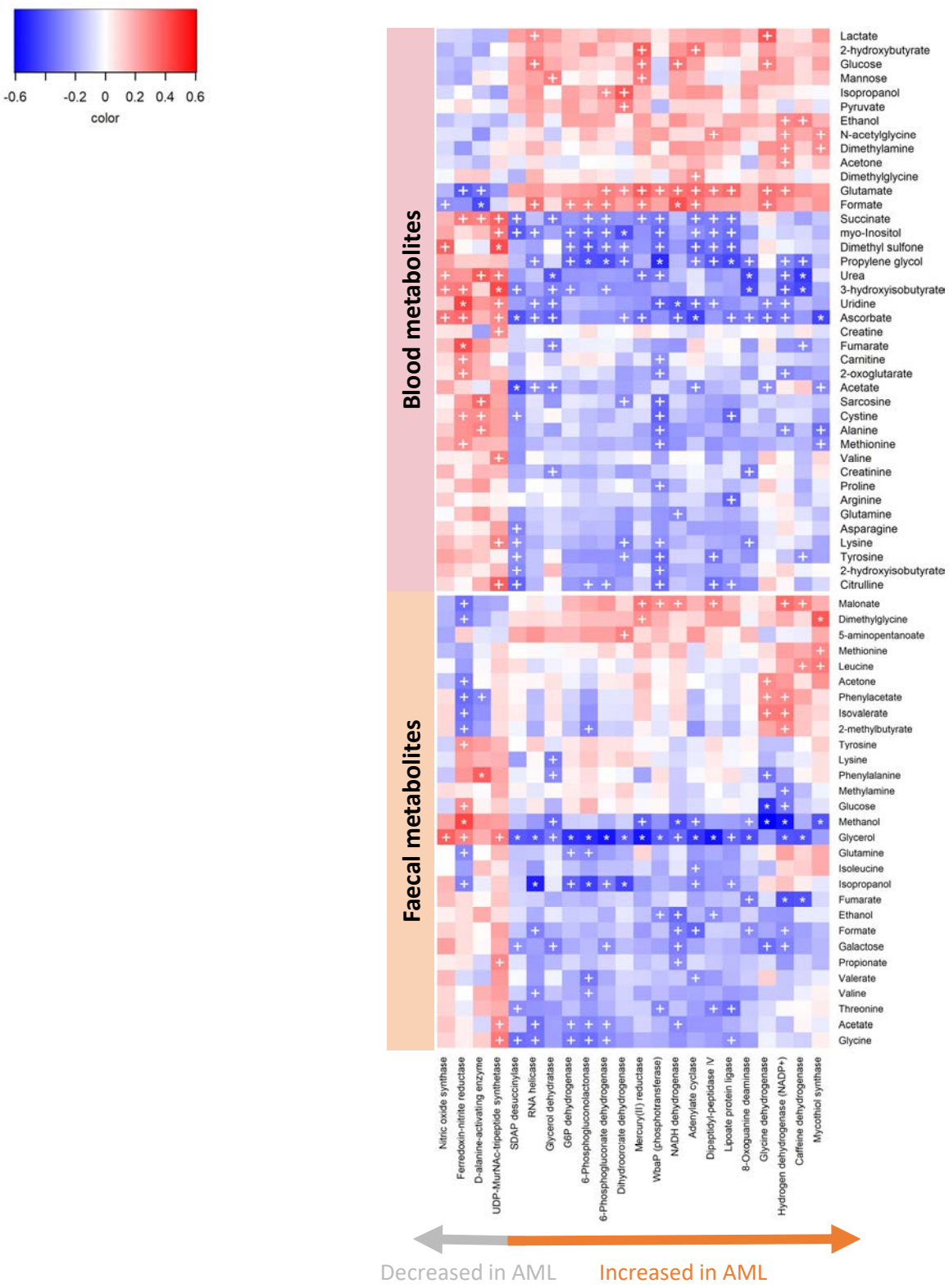



**Figure S15. Correlations between blood and faecal metabolites and altered EC enzyme functions.** Spearman correlations. Metabolites with at least one correlation with an EC enzyme function are present. Microbial functions are ordered by fold change. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1.

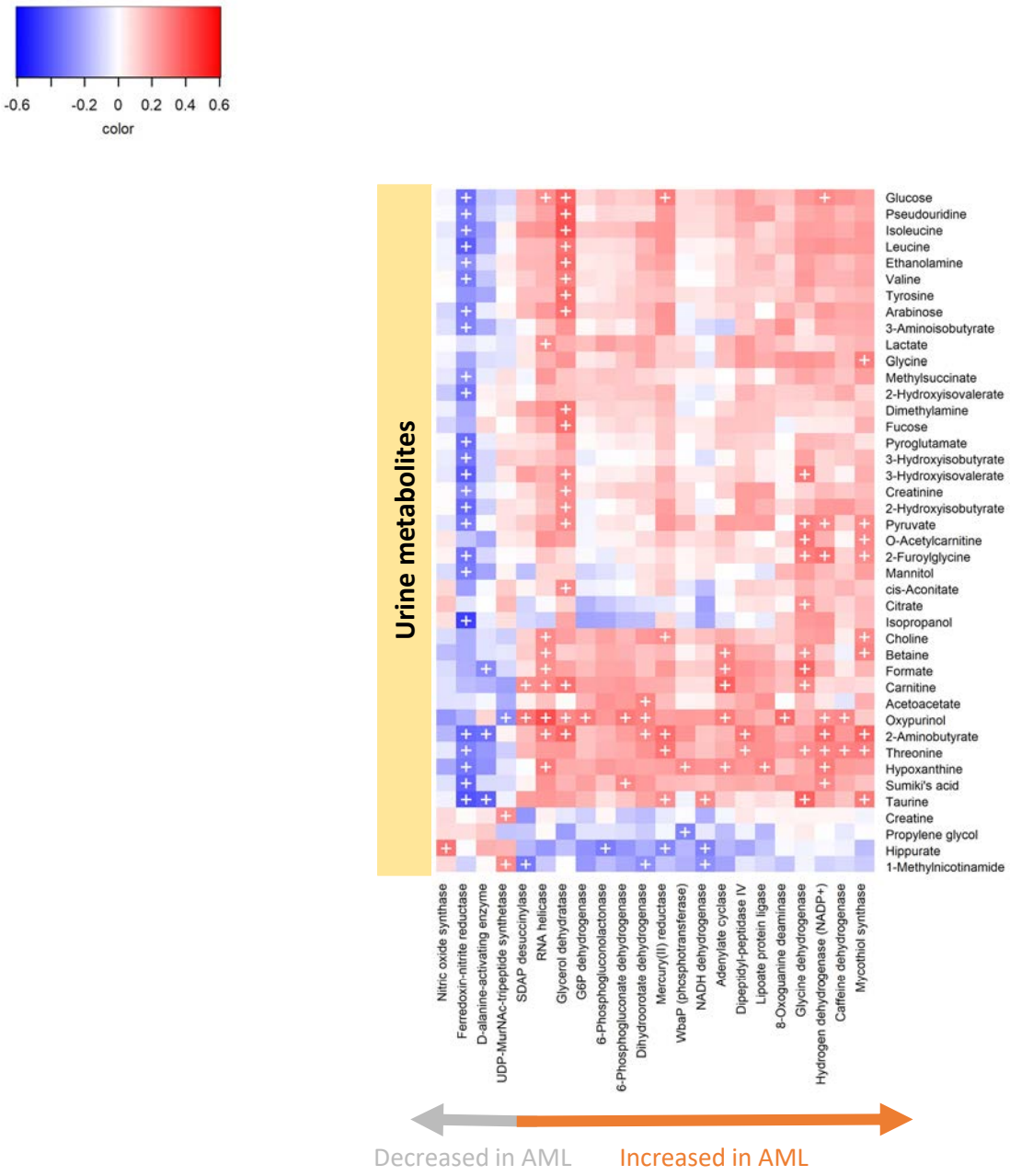



**Figure S16. Correlations between urine metabolites and altered EC enzyme functions.**
Spearman correlations. Metabolites with at least one correlation with an EC enzyme function are present. Microbial functions are ordered by fold change. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1.

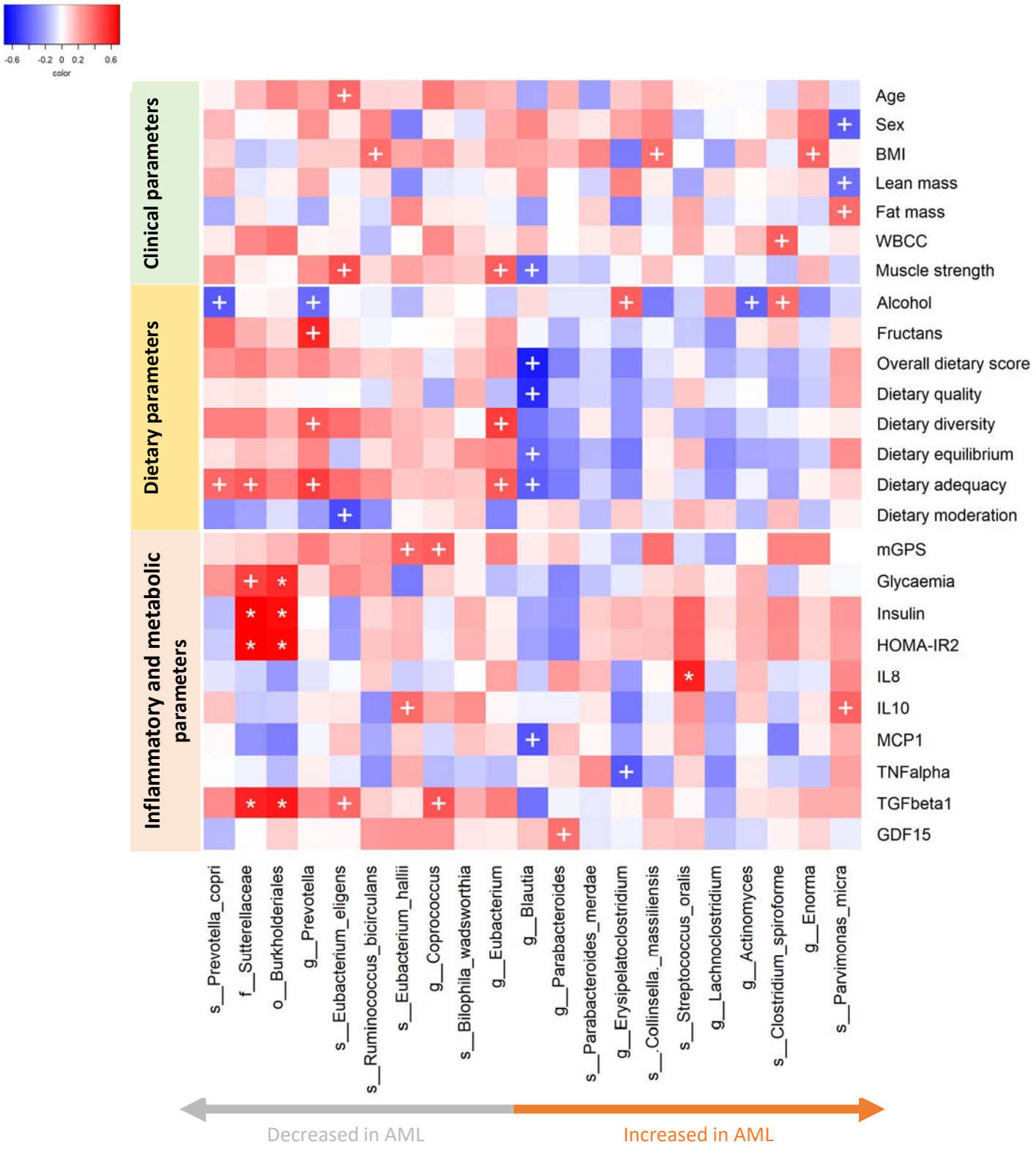



**Figure S17. Correlations between clinical, dietary, inflammatory and metabolic parameters and the top altered bacteria in AML patients exclusively.**

Spearman correlations. Microbial taxa are ordered by fold change. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1. Parameters with at least one correlation with an altered taxon are present. BMI: body mass index; WBCC: white blood cell count; mGPS: modified Glasgow prognostic score; HOMA-IR2: second homeostatic model assessment for insulin resistance; IL8: interleukin-8; IL10: interleukin-10; MCP1: monocyte chemoattractant protein 1; TNFα: tumor necrosis factor alpha-1; TGFβ1: transforming growth factor beta-1; GDF15: growth differentiation factor 15.

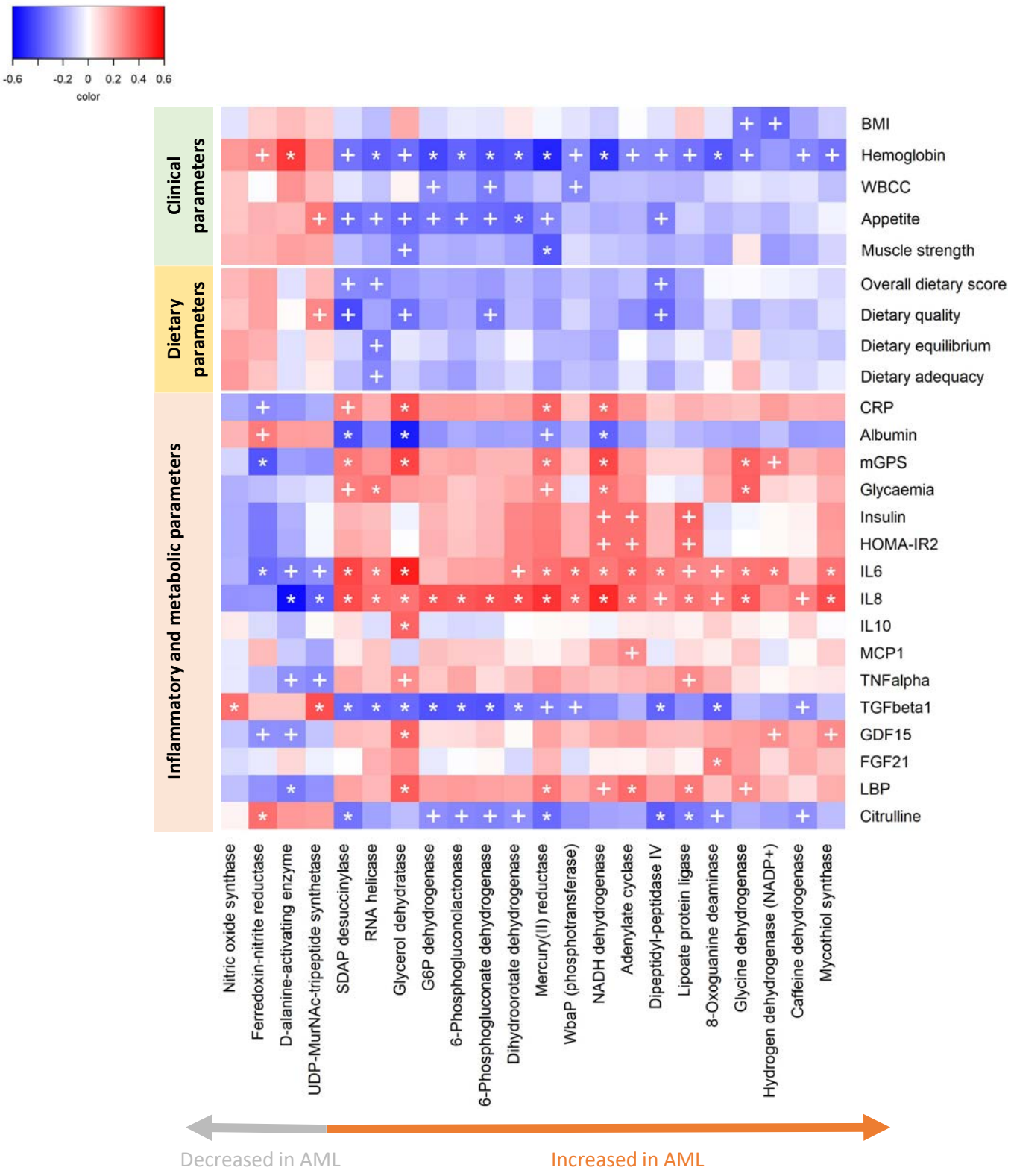



**Figure S18. Correlations between clinical, dietary, inflammatory and metabolic parameters and altered EC enzyme functions in CT subjects and AML patients.**

Spearman correlations. Microbial functions are ordered by fold change. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1. Parameters with at least one correlation with an altered taxon are present. BMI: body mass index; WBCC: white blood cell count; appetite (SNAQ score); CRP: C-reactive protein; mGPS: modified Glasgow prognostic score; HOMA-IR2: second homeostatic model assessment for insulin resistance; IL6: interleukin-6; IL8: interleukin-8; IL10: interleukin-10; MCP1: monocyte chemoattractant protein 1; TNFα: tumor necrosis factor alpha-1; TGFβ1: transforming growth factor beta-1; GDF15: growth differentiation factor 15; FGF21: fibroblast growth factor 21; LBP: lipopolysaccharide binding protein.

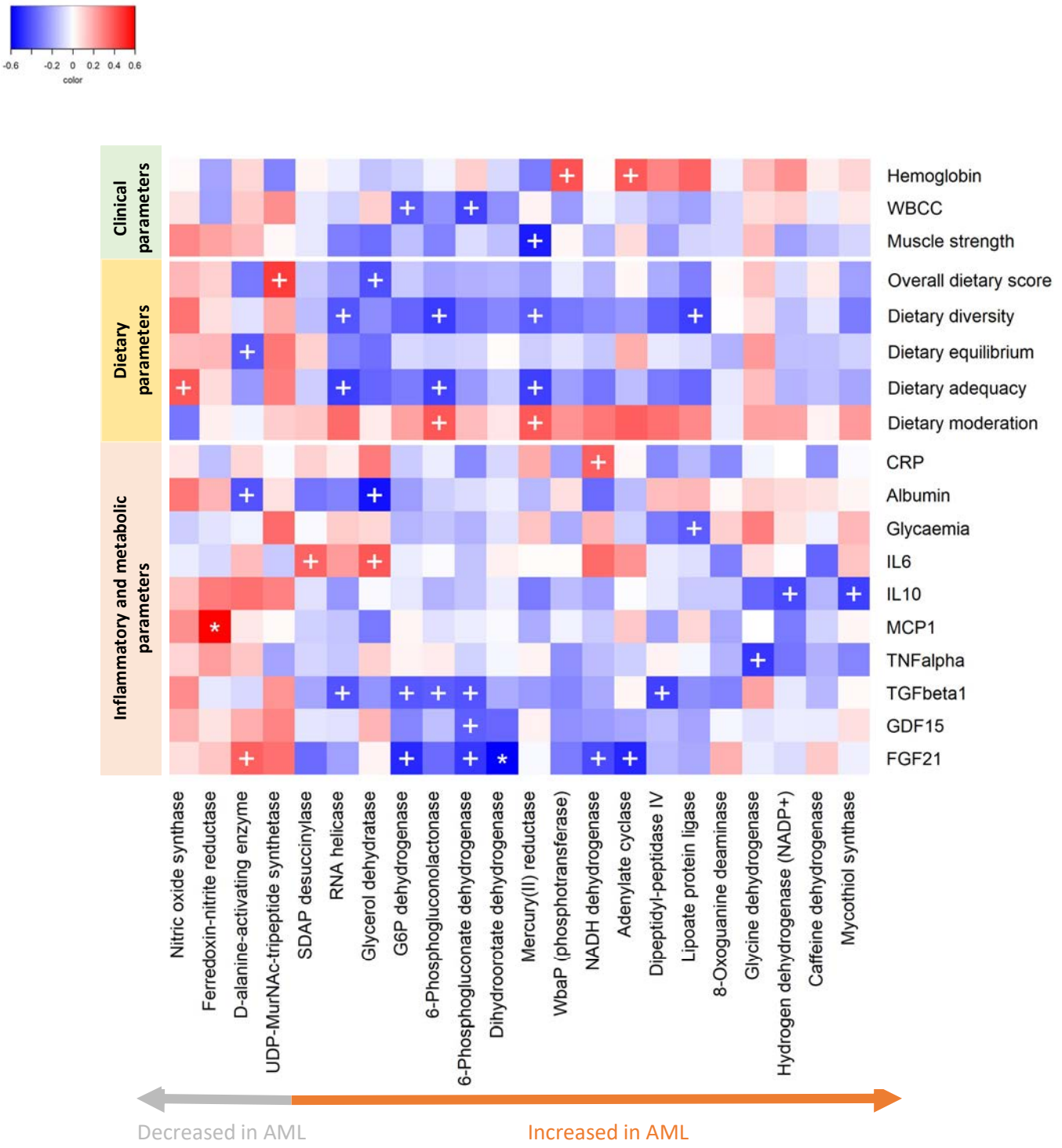



**Figure S19. Correlations between clinical, dietary, inflammatory and metabolic parameters and altered EC enzyme functions in AML patients exclusively.**

Spearman correlations. Microbial functions are ordered by fold change. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1. Parameters with at least one correlation with an altered taxon are present. WBCC: white blood cell count; CRP: C-reactive protein; IL6: interleukin-6; IL10: interleukin-10; MCP1: monocyte chemoattractant protein 1; TNFα: tumor necrosis factor alpha-1; TGFβ1: transforming growth factor beta-1; GDF15: growth differentiation factor 15; FGF21: fibroblast growth factor 21.

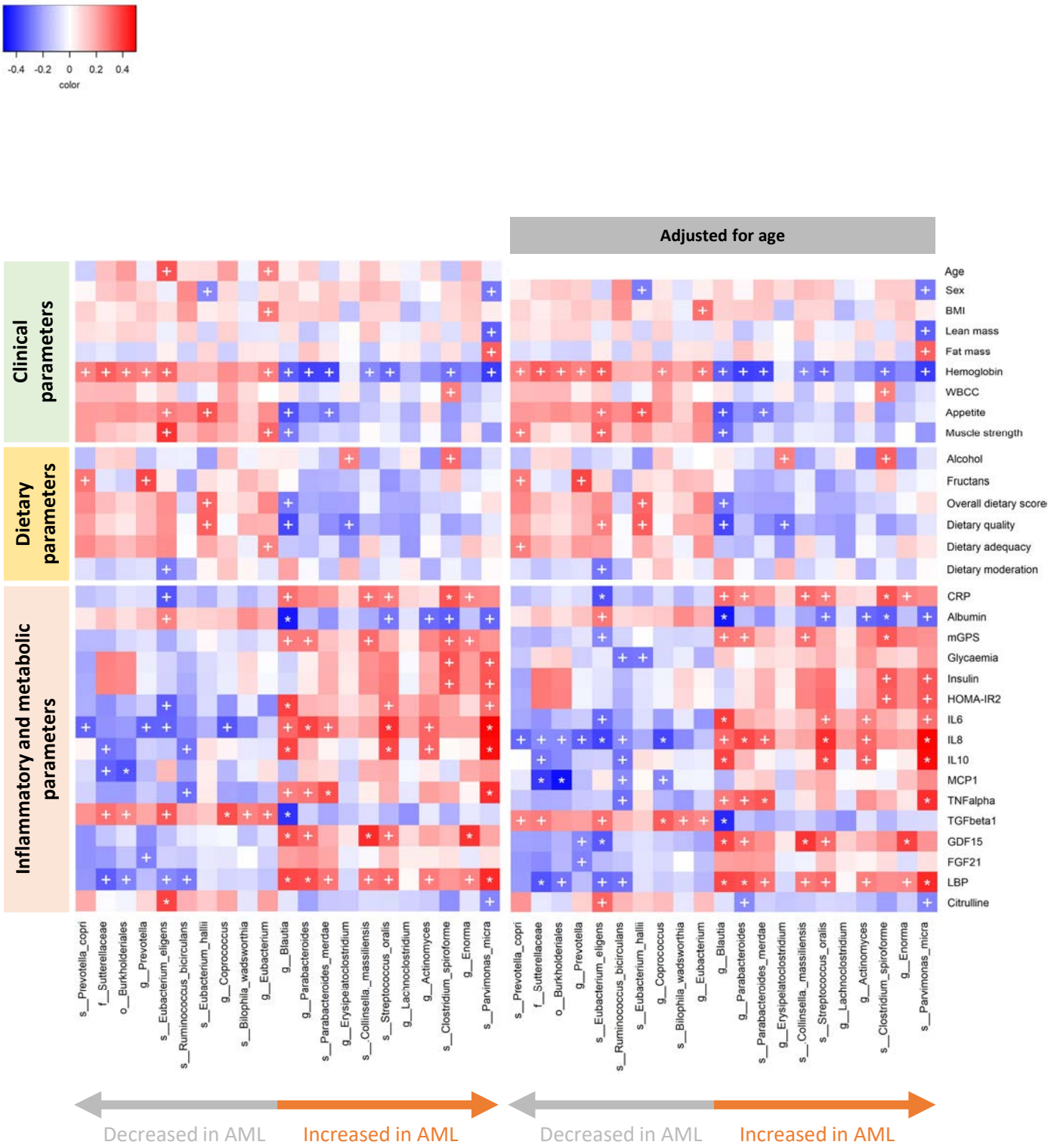



**Figure S20. Correlations between clinical, dietary, inflammatory and metabolic parameters and the top altered bacteria.** Spearman correlations (left) and partial Spearman rank-based correlations (pSRBC) adjusted for age (right) for the whole cohort (AML group and CT group). Metabolites with at least one correlation with an altered taxon are present. '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1.

BMI: body mass index; WBCC: white blood cell count; appetite (SNAQ score); CRP: C-reactive protein; mGPS: modified Glasgow prognostic score; HOMA-IR2: second homeostatic model assessment for insulin resistance; IL6: interleukin-6; IL8: interleukin-8; IL10: interleukin-10; MCP1: monocyte chemoattractant protein 1; TNFα: tumor necrosis factor alpha-1; TGFβ1: transforming growth factor beta-1; GDF15: growth differentiation factor 15; FGF21: fibroblast growth factor 21; LBP: lipopolysaccharide binding protein.

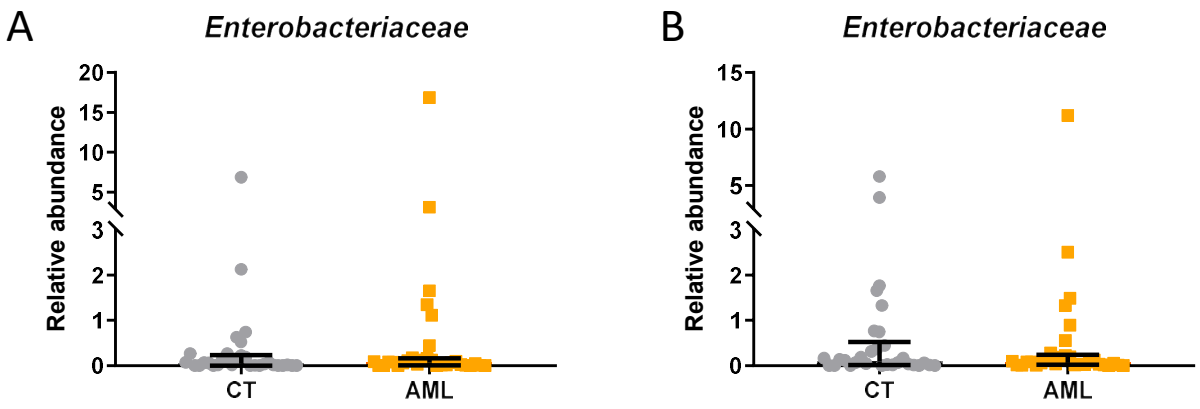



**Figure S21. *Enterobacteriaceae* family levels are not different in AML patients compared to CT subjects**. A) Results obtained using shotgun metagenomics. B) Results obtained using 16S rRNA gene sequencing. Mann-Whitney U-tests with an FDR correction were applied. AML in orange *vs.* CT in grey.

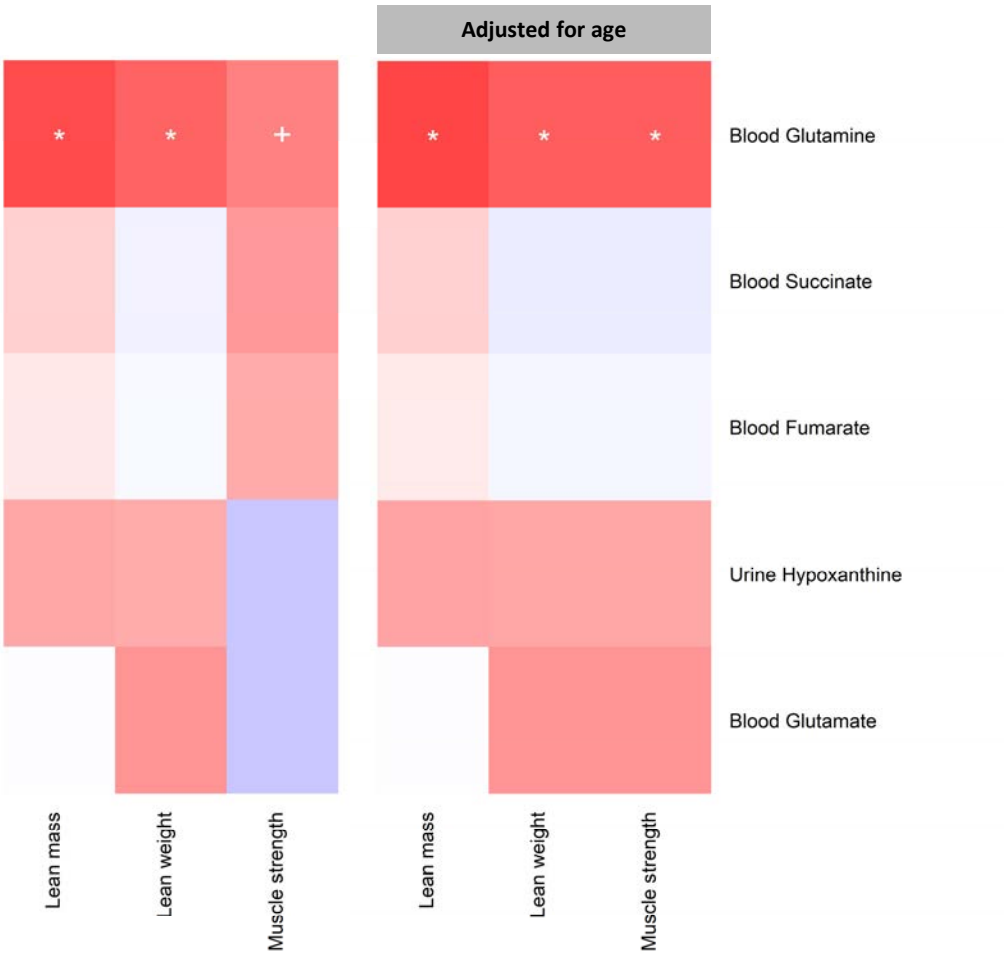



**Figure S22. Correlations between lean mass, lean weight, muscle strength and significantly changed metabolites between AML and CT individuals involved in purine nucleotide metabolism and intense metabolic stress.**

Spearman correlations (left) and partial Spearman rank-based correlations (pSRBC) adjusted for age (right). '+' symbolizes a p-value < 0.05 and '*' symbolizes an FDR-corrected q-value < 0.1.

**Table S1. Drugs and Food Supplements.**

Drugs are grouped by category according to the Belgian classification (CBIP: *Centre Belge d'Information Pharmacothérapeutique*). Only drug categories taken regularly by more than 3 patients in the whole cohort are listed. Less than 3 patients report to take supplementation of amino acids, plants and probiotics. Significance was tested using Fisher's exact test.

| Drug category | CT (n = 30) | AML (n = 30) | Significance |
|---|---|---|---|
| Hypertension (CBIP 1.1) | 8 | 6 | *ns* |
| Adrenergic beta-antagonists (CBIP 1.5) | 5 | 3 | *ns* |
| Calcium channel clockers (CBIP 1.6) | 4 | 3 | *ns* |
| Drugs acting on the renin-angiotensin system (CBIP 1.7) | 5 | 4 | *ns* |
| Hypolipidemic agents (CBIP 1.12) | 6 | 4 | *ns* |
| Antithrombotic agents(CBIP 2.1) | 3 | 8 | *ns* |
| Gastric and duodenal pathologies (CBIP 3.1) | 2 | 5 | *ns* |
| Laxatives (CBIP 3.5) | 1 | 3 | *ns* |
| Contraception (CBIP 6.2) | 2 | 3 | *ns* |
| Prostate benign hypertrophy (CBIP 7.2) | 1 | 3 | *ns* |
| Gout (CBIP 9.3) | 1 | 4 | *ns* |
| Osteoporosis and Paget disease (CBIP 9.5) | 4 | 2 | *ns* |
| Hypnotics, sedatives, anxiolytics (CBIP 10.1) | 2 | 2 | *ns* |
| Antidepressants (CBIP 10.3) | 5 | 4 | *ns* |
| Allergies (CBIP 12.4) | 2 | 2 | *ns* |
| Minerals (CBIP 14.1) | 2 | 2 | *ns* |
| Vitamins (CBIP 14.1) | 7 | 4 | *ns* |
| Rhinitis and sinusitis (CBIP 17.3) | 3 | 2 | *ns* |

NB: Anti-Bacterial agents were taken by 2 CT subjects and 1 AML patient between day -30 and day -90 before inclusion.

**Table S2. Top altered bacteria in AML patients.**

Top 21 bacteria selected based on top bacteria from untargeted metagenomics analyses. Bacteria were selected based on p-values from Mann-Whitney U-test on raw data (MW), Mann-Whitney U-test on centered log-ratio data (MW-CLR), and ALDEx2. Results of targeted metagenomics (16S rRNA gene sequencing, Mann-Whitney U-test, MW) are also mentioned (p-value). IQR: interquartile range.

| | Shotgun metagenomics | | | | | | | | 16S rRNA gene sequencing | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Median | | IQR | | MW | | MW-CLR | ALDEx2 | MW | |
| | CT | AML | CT | AML | p-value | q-value | p-value | p-value | p-value | q-value |
| s__Parvimonas_micra | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.067 | 0.004 | 0.395 | ND | ND |
| s__Eubacterium_eligens | 1.363 | 0.364 | 2.370 | 1.276 | 0.001 | 0.074 | 0.000 | 0.000 | ND | ND |
| g__Parabacteroides | 1.235 | 2.137 | 1.011 | 1.716 | 0.002 | 0.075 | 0.061 | 0.020 | 0.006 | 0.112 |
| g__Actinomyces | 0.022 | 0.052 | 0.028 | 0.099 | 0.003 | 0.076 | 0.001 | 0.009 | 0.001 | 0.046 |
| g__Blautia | 1.918 | 3.144 | 1.602 | 1.772 | 0.004 | 0.094 | 0.063 | 0.032 | 0.723 | 0.952 |
| s__Streptococcus_oralis | 0.000 | 0.006 | 0.002 | 0.018 | 0.005 | 0.106 | 0.012 | 0.104 | ND | ND |
| s__Clostridium_spiroforme | 0.000 | 0.008 | 0.003 | 0.085 | 0.005 | 0.106 | 0.007 | 0.045 | ND | ND |
| g__Prevotella | 3.388 | 0.224 | 19.569 | 2.856 | 0.006 | 0.109 | 0.005 | 0.007 | 0.004 | 0.100 |
| g__Coprococcus | 2.371 | 1.223 | 2.598 | 1.796 | 0.006 | 0.109 | 0.000 | 0.001 | 0.001 | 0.046 |
| s__Prevotella_copri | 0.030 | 0.000 | 10.689 | 0.011 | 0.008 | 0.131 | 0.024 | 0.027 | ND | ND |
| s__Parabacteroides_merdae | 0.340 | 0.700 | 0.659 | 0.937 | 0.010 | 0.163 | 0.010 | 0.011 | ND | ND |
| s__Eubacterium_hallii | 1.461 | 0.955 | 1.295 | 1.387 | 0.015 | 0.186 | 0.013 | 0.011 | ND | ND |
| g__Lachnoclostridium | 0.010 | 0.090 | 0.055 | 0.215 | 0.016 | 0.186 | 0.051 | 0.020 | ND | ND |
| f__Sutterellaceae | 0.059 | 0.020 | 0.530 | 0.114 | 0.016 | 0.186 | 0.004 | 0.013 | 0.030 | 0.236 |
| g__Eubacterium | 6.290 | 4.124 | 6.168 | 2.901 | 0.019 | 0.201 | 0.027 | 0.032 | ND | ND |
| s__.Collinsella._massiliensis | 0.000 | 0.005 | 0.001 | 0.041 | 0.019 | 0.201 | 0.008 | 0.069 | ND | ND |
| g__Erysipelatoclostridium | 0.013 | 0.105 | 0.046 | 0.357 | 0.021 | 0.206 | 0.032 | 0.035 | ND | ND |
| g__Enorma | 0.000 | 0.016 | 0.002 | 0.056 | 0.028 | 0.227 | 0.010 | 0.089 | ND | ND |
| s__Ruminococcus_bicirculans | 0.416 | 0.041 | 1.780 | 0.392 | 0.033 | 0.227 | 0.026 | 0.034 | ND | ND |
| s__Bilophila_wadsworthia | 0.053 | 0.024 | 0.071 | 0.064 | 0.036 | 0.227 | 0.007 | 0.056 | ND | ND |
| o__Burkholderiales | 0.059 | 0.031 | 0.534 | 0.144 | 0.064 | 0.301 | 0.014 | 0.042 | 0.011 | 0.134 |

**Table S3. List of the 22 EC enzyme functions altered in AML patients compared to CT, as assessed using MaAsLin2.**

For sake of clarity, abbreviated names are used throughout the paper. Correspondence between EC nomenclature, full function names and abbreviated names are included here.

| EC function | Full name | Abbreviated name |
|---|---|---|
| **EC 1.14.13.39** | Nitric-oxide synthase (NADPH) | Nitric oxide synthase |
| **EC 1.7.7.1** | Ferredoxin-nitrite reductase | Ferredoxin nitrite reductase |
| **EC 6.1.1.13** | D-alanine-poly(phosphoribitol) ligase | D-alanine-activating enzyme |
| **EC 6.3.2.13** | UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase | UDP-MurNAc-tripeptide synthetase |
| **EC 3.5.1.18** | Succinyl-diaminopimelate desuccinylase | SDAP desuccinylase |
| **EC 3.6.4.13** | RNA helicase | RNA helicase |
| **EC 4.2.1.30** | Glycerol dehydratase | Glycerol dehydratase |
| **EC 1.1.1.49** | Glucose-6-phosphate dehydrogenase (NADP+) | G6P dehydrogenase |
| **EC 3.1.1.31** | 6-Phosphogluconolactonase | 6-Phosphogluconolactonase |
| **EC 1.1.1.44** | Phosphogluconate dehydrogenase (NADP dependent, decarboxylating) | 6-Phosphogluconate dehydrogenase |
| **EC 1.3.98.1** | Dihydroorotate dehydrogenase (fumarate) | Dihydroorotate dehydrogenase |
| **EC 1.16.1.1** | Mercury(II) reductase | Mercury(II) reductase |
| **EC 2.7.8.6** | Undecaprenyl-phosphate galactose phosphotransferase | WbaP (phosphotransferase) |
| **EC 1.6.99.3** | NADH dehydrogenase | NADH dehydrogenase |
| **EC 4.6.1.1** | Adenylate cyclase | Adenylate cyclase |
| **EC 3.4.14.5** | Dipeptidyl-peptidase IV | Dipeptidyl-peptidase IV |
| **EC 2.7.7.63** | Lipoate protein ligase | Lipoate protein ligase |
| **EC 3.5.4.32** | 8-Oxoguanine deaminase | 8-Oxoguanine deaminase |
| **EC 1.4.99.5** | Glycine dehydrogenase (cyanide-forming) | Glycine dehydrogenase |
| **EC 1.12.1.3** | Hydrogen dehydrogenase (NADP+) | Hydrogen dehydrogenase (NADP+) |
| **EC 1.17.5.2** | Caffeine dehydrogenase | Caffeine dehydrogenase |
| **EC 2.3.1.189** | Mycothiol synthase | Mycothiol synthase |

**Table S4. Median, interquartile range (IQR), p-value of Mann-Whitney U-test and q-value after false discovery rate correction for all metabolites in the three compartments.** Imputation was not performed on maleate and 3-phenylpropionate due to more than 40% missing values in one group. For those metabolites, the differences between CT and AML was tested using a Fisher test. No q-value is therefore reported for those metabolites.

| Metabolite | Faeces | | | | | | Blood | | | | | | Urine | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Median | | IQR | | Mann-Whitney | | Median | | IQR | | Mann-Whitney | | Median | | IQR | | Mann-Whitney | |
| | CT | AML | CT | AML | p-value | q-value | CT | AML | CT | AML | p-value | q-value | CT | AML | CT | AML | p-value | q-value |
| Acetate | 20.995 | 17.499 | 10.738 | 12.570 | 0.485 | 0.635 | 0.033 | 0.024 | 0.015 | 0.019 | 0.010 | 0.053 | NA | NA | NA | NA | NA | NA |
| Acetoacetate | NA | NA | NA | NA | NA | NA | 0.020 | 0.021 | 0.024 | 0.023 | 0.294 | 0.438 | 0.185 | 0.241 | 0.195 | 0.536 | 0.449 | 0.625 |
| Acetone | 0.030 | 0.024 | 0.022 | 0.027 | 0.888 | 0.888 | 0.008 | 0.011 | 0.004 | 0.013 | 0.240 | 0.382 | 0.112 | 0.116 | 0.047 | 0.130 | 0.663 | 0.772 |
| Alanine | 0.273 | 0.192 | 0.130 | 0.166 | 0.181 | 0.382 | 0.258 | 0.208 | 0.076 | 0.094 | 0.010 | 0.053 | 0.929 | 0.822 | 0.965 | 0.961 | 0.467 | 0.636 |
| Allantoin | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.400 | 0.369 | 0.339 | 0.383 | 0.631 | 0.772 |
| Arabinose | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.147 | 1.230 | 0.869 | 1.107 | 0.260 | 0.545 |
| Arginine | NA | NA | NA | NA | NA | NA | 0.051 | 0.042 | 0.018 | 0.020 | 0.171 | 0.307 | NA | NA | NA | NA | NA | NA |
| Ascorbate | NA | NA | NA | NA | NA | NA | 0.035 | 0.011 | 0.013 | 0.022 | 0.000 | 0.000 | NA | NA | NA | NA | NA | NA |
| Asparagine | NA | NA | NA | NA | NA | NA | 0.029 | 0.024 | 0.004 | 0.012 | 0.029 | 0.089 | NA | NA | NA | NA | NA | NA |
| Aspartate | 0.321 | 0.268 | 0.164 | 0.178 | 0.644 | 0.730 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Betaine | NA | NA | NA | NA | NA | NA | 0.023 | 0.024 | 0.011 | 0.012 | 1.000 | 1.000 | 0.283 | 0.400 | 0.484 | 0.589 | 0.209 | 0.545 |
| Butyrate | 4.785 | 3.568 | 2.982 | 3.237 | 0.374 | 0.578 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Carnitine | NA | NA | NA | NA | NA | NA | 0.027 | 0.025 | 0.008 | 0.012 | 0.395 | 0.507 | 0.138 | 0.477 | 0.491 | 0.656 | 0.003 | 0.069 |
| Choline | NA | NA | NA | NA | NA | NA | 0.005 | 0.005 | 0.002 | 0.003 | 0.668 | 0.730 | 0.165 | 0.210 | 0.138 | 0.168 | 0.301 | 0.545 |
| cis-Aconitate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.003 | 0.986 | 0.853 | 0.781 | 0.676 | 0.772 |
| Citrate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 11.271 | 10.451 | 9.905 | 10.020 | 0.423 | 0.616 |
| Citrulline | NA | NA | NA | NA | NA | NA | 0.020 | 0.016 | 0.008 | 0.008 | 0.019 | 0.073 | NA | NA | NA | NA | NA | NA |
| Creatine | NA | NA | NA | NA | NA | NA | 0.017 | 0.018 | 0.012 | 0.017 | 0.464 | 0.571 | 0.647 | 0.269 | 1.829 | 0.519 | 0.248 | 0.545 |
| Creatine phosphate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.329 | 0.284 | 0.405 | 0.324 | 0.242 | 0.545 |
| Creatinine | NA | NA | NA | NA | NA | NA | 0.055 | 0.051 | 0.014 | 0.014 | 0.188 | 0.327 | 48.632 | 46.992 | 37.005 | 46.704 | 0.562 | 0.709 |
| Cystine | NA | NA | NA | NA | NA | NA | 0.044 | 0.031 | 0.017 | 0.028 | 0.044 | 0.117 | NA | NA | NA | NA | NA | NA |
| Dimethylglycine | 0.005 | 0.004 | 0.005 | 0.013 | 0.183 | 0.382 | 0.002 | 0.002 | 0.001 | 0.001 | 0.141 | 0.267 | 0.206 | 0.196 | 0.146 | 0.232 | 0.654 | 0.772 |
| Dimethyl sulfone | NA | NA | NA | NA | NA | NA | 0.007 | 0.006 | 0.004 | 0.004 | 0.109 | 0.229 | NA | NA | NA | NA | NA | NA |
| Dimethylamine | NA | NA | NA | NA | NA | NA | 0.001 | 0.001 | 0.000 | 0.001 | 0.027 | 0.088 | 2.409 | 2.207 | 1.556 | 3.111 | 0.350 | 0.596 |
| Ethanol | 0.146 | 0.032 | 0.081 | 0.094 | 0.011 | 0.098 | 0.002 | 0.003 | 0.001 | 0.004 | 0.006 | 0.038 | NA | NA | NA | NA | NA | NA |
| Ethanolamine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.613 | 1.755 | 1.532 | 1.819 | 0.485 | 0.647 |
| Formate | 0.006 | 0.002 | 0.008 | 0.005 | 0.040 | 0.162 | 0.009 | 0.014 | 0.004 | 0.007 | 0.000 | 0.000 | 1.037 | 1.633 | 0.871 | 1.608 | 0.004 | 0.069 |
| Fucose | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.578 | 0.622 | 0.540 | 0.542 | 0.382 | 0.596 |
| Fumarate | 0.035 | 0.023 | 0.035 | 0.031 | 0.363 | 0.578 | 0.000 | 0.000 | 0.000 | 0.000 | 0.034 | 0.101 | 0.015 | 0.023 | 0.028 | 0.021 | 0.404 | 0.615 |
| Galactose | 0.131 | 0.053 | 0.077 | 0.076 | 0.002 | 0.023 | NA | NA | NA | NA | NA | NA | 0.630 | 0.824 | 0.698 | 0.598 | 0.889 | 0.903 |
| Glucose | 0.694 | 0.257 | 0.523 | 0.308 | 0.043 | 0.162 | 4.018 | 4.323 | 0.621 | 1.785 | 0.020 | 0.073 | 1.419 | 1.897 | 1.189 | 1.848 | 0.028 | 0.172 |
| Glutamate | 1.451 | 1.039 | 0.574 | 0.911 | 0.542 | 0.658 | 0.015 | 0.032 | 0.011 | 0.013 | 0.000 | 0.000 | NA | NA | NA | NA | NA | NA |
| Glutamine | 0.264 | 0.223 | 0.163 | 0.193 | 0.485 | 0.635 | 0.355 | 0.318 | 0.052 | 0.090 | 0.022 | 0.075 | NA | NA | NA | NA | NA | NA |
| Glycerol | 0.927 | 0.087 | 1.120 | 0.215 | 0.000 | 0.003 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Glycine | 0.134 | 0.090 | 0.072 | 0.060 | 0.072 | 0.230 | NA | NA | NA | NA | NA | NA | 4.035 | 5.712 | 4.188 | 5.537 | 0.307 | 0.545 |
| Hippurate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 14.958 | 7.931 | 13.200 | 10.889 | 0.011 | 0.112 |
| Hypoxanthine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.144 | 0.269 | 0.208 | 0.380 | 0.004 | 0.069 |
| Isobutyrate | 0.829 | 0.728 | 0.566 | 0.920 | 0.213 | 0.382 | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.136 | NA | NA | NA | NA | NA | NA |
| Isoleucine | 0.097 | 0.060 | 0.081 | 0.063 | 0.228 | 0.388 | 0.044 | 0.041 | 0.014 | 0.019 | 0.249 | 0.386 | 0.051 | 0.063 | 0.036 | 0.069 | 0.041 | 0.172 |
| Isopropanol | 0.029 | 0.014 | 0.033 | 0.021 | 0.756 | 0.829 | 0.000 | 0.001 | 0.000 | 0.001 | 0.085 | 0.186 | 0.033 | 0.035 | 0.030 | 0.036 | 0.701 | 0.787 |
| Isovalerate | 0.567 | 0.557 | 0.409 | 0.897 | 0.150 | 0.376 | NA | NA | NA | NA | NA | NA | 0.034 | 0.030 | 0.032 | 0.037 | 0.929 | 0.929 |
| Lactate | NA | NA | NA | NA | NA | NA | 0.761 | 1.108 | 0.380 | 0.425 | 0.000 | 0.003 | 0.301 | 0.318 | 0.292 | 0.258 | 0.423 | 0.616 |
| Leucine | 0.132 | 0.101 | 0.092 | 0.083 | 0.802 | 0.852 | 0.080 | 0.081 | 0.021 | 0.038 | 0.918 | 0.933 | 0.088 | 0.115 | 0.053 | 0.130 | 0.041 | 0.172 |
| Lysine | 0.130 | 0.079 | 0.102 | 0.071 | 0.081 | 0.230 | 0.115 | 0.101 | 0.033 | 0.046 | 0.075 | 0.170 | 0.317 | 0.311 | 0.382 | 0.860 | 0.665 | 0.772 |
| Maleate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.020 | 0.029 | 0.016 | 0.023 | 0.000 | NA |
| Malonate | 0.155 | 0.135 | 0.090 | 0.166 | 0.155 | 0.376 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Mannitol | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.376 | 2.240 | 2.284 | 3.050 | 0.155 | 0.450 |
| Mannose | NA | NA | NA | NA | NA | NA | 0.043 | 0.059 | 0.012 | 0.036 | 0.012 | 0.053 | NA | NA | NA | NA | NA | NA |
| Methanol | 0.218 | 0.068 | 0.124 | 0.067 | 0.000 | 0.003 | NA | NA | NA | NA | NA | NA | 0.915 | 0.784 | 1.106 | 0.791 | 0.449 | 0.625 |
| Methionine | 0.067 | 0.062 | 0.039 | 0.056 | 0.595 | 0.697 | 0.018 | 0.016 | 0.005 | 0.006 | 0.018 | 0.073 | NA | NA | NA | NA | NA | NA |
| Methylamine | 0.059 | 0.044 | 0.036 | 0.043 | 0.520 | 0.655 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Methylguanidine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.153 | 0.102 | 0.265 | 0.163 | 0.300 | 0.545 |
| Methylsuccinate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.040 | 0.055 | 0.049 | 0.068 | 0.176 | 0.490 |
| myo-Inositol | NA | NA | NA | NA | NA | NA | 0.021 | 0.020 | 0.008 | 0.009 | 0.379 | 0.507 | NA | NA | NA | NA | NA | NA |
| N-Acetylglycine | NA | NA | NA | NA | NA | NA | 0.001 | 0.002 | 0.001 | 0.002 | 0.158 | 0.290 | NA | NA | NA | NA | NA | NA |
| O-Acetylcarnitine | NA | NA | NA | NA | NA | NA | 0.005 | 0.005 | 0.002 | 0.004 | 0.351 | 0.487 | 0.074 | 0.107 | 0.073 | 0.113 | 0.146 | 0.446 |
| Ornithine | NA | NA | NA | NA | NA | NA | 0.036 | 0.040 | 0.011 | 0.019 | 0.530 | 0.638 | NA | NA | NA | NA | NA | NA |
| Oxypurinol | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.043 | 2.105 | 2.184 | 29.182 | 0.286 | 0.545 |
| Pantothenate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.073 | 0.076 | 0.056 | 0.106 | 0.819 | 0.873 |
| Phenylacetate | 0.231 | 0.224 | 0.134 | 0.265 | 0.040 | 0.162 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Phenylalanine | 0.060 | 0.034 | 0.044 | 0.038 | 0.020 | 0.135 | 0.036 | 0.042 | 0.009 | 0.013 | 0.057 | 0.136 | NA | NA | NA | NA | NA | NA |
| Proline | NA | NA | NA | NA | NA | NA | 0.130 | 0.109 | 0.060 | 0.039 | 0.198 | 0.334 | NA | NA | NA | NA | NA | NA |
| Propionate | 5.949 | 4.822 | 3.060 | 3.917 | 0.432 | 0.635 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Propylene glycol | NA | NA | NA | NA | NA | NA | 0.001 | 0.001 | 0.002 | 0.002 | 0.387 | 0.507 | 0.108 | 0.095 | 0.158 | 0.118 | 0.854 | 0.882 |
| Pseudouridine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.683 | 0.847 | 0.471 | 0.934 | 0.037 | 0.172 |
| Pyroglutamate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.088 | 1.239 | 0.610 | 1.083 | 0.382 | 0.596 |
| Pyruvate | NA | NA | NA | NA | NA | NA | 0.056 | 0.060 | 0.029 | 0.020 | 0.343 | 0.487 | 0.080 | 0.096 | 0.075 | 0.155 | 0.530 | 0.678 |
| Sarcosine | NA | NA | NA | NA | NA | NA | 0.001 | 0.001 | 0.001 | 0.001 | 0.049 | 0.127 | NA | NA | NA | NA | NA | NA |
| Serine | NA | NA | NA | NA | NA | NA | 0.079 | 0.083 | 0.016 | 0.028 | 0.762 | 0.803 | NA | NA | NA | NA | NA | NA |
| sn-Glycero-3-phosphocholin | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.219 | 0.237 | 0.177 | 0.237 | 0.286 | 0.545 |
| Succinate | NA | NA | NA | NA | NA | NA | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.007 | NA | NA | NA | NA | NA | NA |
| Sumiki's acid | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.058 | 0.121 | 0.115 | 0.300 | 0.030 | 0.172 |
| Taurine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.479 | 5.266 | 3.995 | 8.152 | 0.001 | 0.037 |
| Threonine | 0.109 | 0.062 | 0.063 | 0.064 | 0.030 | 0.162 | 0.097 | 0.079 | 0.026 | 0.043 | 0.038 | 0.108 | 0.333 | 0.548 | 0.290 | 0.480 | 0.018 | 0.141 |
| Trigonelline | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1.309 | 0.692 | 2.321 | 1.814 | 0.014 | 0.125 |
| Trimethylamine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.078 | 0.069 | 0.096 | 0.073 | 0.523 | 0.678 |
| Tryptophan | NA | NA | NA | NA | NA | NA | 0.005 | 0.005 | 0.002 | 0.002 | 0.646 | 0.730 | NA | NA | NA | NA | NA | NA |
| Tyrosine | 0.072 | 0.051 | 0.053 | 0.036 | 0.196 | 0.382 | 0.052 | 0.052 | 0.015 | 0.025 | 0.564 | 0.666 | 0.388 | 0.663 | 0.345 | 0.682 | 0.028 | 0.172 |
| Uracil | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.205 | 0.136 | 0.186 | 0.136 | 0.034 | 0.172 |
| Urea | NA | NA | NA | NA | NA | NA | 0.772 | 0.637 | 0.349 | 0.209 | 0.011 | 0.053 | 40.891 | 35.550 | 17.906 | 16.420 | 0.119 | 0.382 |
| Uridine | NA | NA | NA | NA | NA | NA | 0.003 | 0.002 | 0.001 | 0.001 | 0.000 | 0.000 | NA | NA | NA | NA | NA | NA |
| Valerate | 0.851 | 0.645 | 0.504 | 0.759 | 0.859 | 0.885 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Valine | 0.164 | 0.104 | 0.101 | 0.064 | 0.076 | 0.230 | 0.163 | 0.165 | 0.032 | 0.055 | 0.728 | 0.781 | 0.133 | 0.164 | 0.122 | 0.252 | 0.095 | 0.337 |
| 1,5-Anhydrosorbitol | NA | NA | NA | NA | NA | NA | 0.085 | 0.070 | 0.043 | 0.030 | 0.654 | 0.730 | NA | NA | NA | NA | NA | NA |
| 1-Methylnicotinamide | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.187 | 0.141 | 0.174 | 0.185 | 0.236 | 0.545 |
| 2-Aminobutyrate | NA | NA | NA | NA | NA | NA | 0.012 | 0.014 | 0.007 | 0.008 | 0.297 | 0.438 | 0.047 | 0.078 | 0.024 | 0.080 | 0.007 | 0.089 |
| 2-Furoylglycine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.121 | 0.259 | 0.398 | 0.875 | 0.236 | 0.545 |
| 2-Hydroxybutyrate | NA | NA | NA | NA | NA | NA | 0.023 | 0.030 | 0.010 | 0.020 | 0.001 | 0.007 | NA | NA | NA | NA | NA | NA |
| 2-Hydroxyisobutyrate | NA | NA | NA | NA | NA | NA | 0.001 | 0.001 | 0.000 | 0.001 | 0.462 | 0.571 | 0.200 | 0.226 | 0.174 | 0.151 | 0.243 | 0.545 |
| 2-Hydroxyisovalerate | NA | NA | NA | NA | NA | NA | 0.005 | 0.006 | 0.002 | 0.006 | 0.214 | 0.351 | 0.023 | 0.029 | 0.026 | 0.038 | 0.363 | 0.596 |
| 2-Methylbutyrate | 0.919 | 0.783 | 0.628 | 1.367 | 0.449 | 0.635 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 2-Oxoglutarate | NA | NA | NA | NA | NA | NA | 0.005 | 0.005 | 0.002 | 0.002 | 0.657 | 0.730 | 0.433 | 0.500 | 0.428 | 0.490 | 0.358 | 0.596 |
| 2-Oxoisocaproate | NA | NA | NA | NA | NA | NA | 0.003 | 0.004 | 0.001 | 0.002 | 0.120 | 0.240 | NA | NA | NA | NA | NA | NA |
| 3-Aminoisobutyrate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.346 | 0.501 | 0.402 | 1.809 | 0.049 | 0.187 |
| 3-Hydroxybutyrate | NA | NA | NA | NA | NA | NA | 0.028 | 0.027 | 0.039 | 0.039 | 0.859 | 0.885 | NA | NA | NA | NA | NA | NA |
| 3-Hydroxyisobutyrate | NA | NA | NA | NA | NA | NA | 0.012 | 0.009 | 0.004 | 0.005 | 0.122 | 0.240 | 0.301 | 0.328 | 0.321 | 0.382 | 0.307 | 0.545 |
| 3-Hydroxyisovalerate | NA | NA | NA | NA | NA | NA | 0.001 | 0.001 | 0.000 | 0.001 | 0.006 | 0.038 | 0.229 | 0.275 | 0.199 | 0.310 | 0.107 | 0.361 |
| 3-Hydroxymandelate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.324 | 0.285 | 0.320 | 0.634 | 0.741 | 0.818 |
| 3-Indoxylsulfate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.995 | 0.937 | 1.212 | 1.173 | 0.854 | 0.882 |
| 3-Methyl-2-oxovalerate | NA | NA | NA | NA | NA | NA | 0.003 | 0.003 | 0.001 | 0.001 | 0.355 | 0.487 | NA | NA | NA | NA | NA | NA |
| 3-Methylxanthine | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.196 | 0.099 | 0.280 | 0.132 | 0.043 | 0.172 |
| 3-Phenylpropionate | 0.090 | 0.061 | 0.060 | 0.036 | 0.007 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 4-Hydroxyphenylacetate | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.484 | 0.538 | 0.406 | 0.478 | 0.786 | 0.853 |
| 5-Aminopentanoate | 0.051 | 0.040 | 0.045 | 0.079 | 0.208 | 0.382 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |

**Table S5. Significantly changed bacterial species in AML patients with high insulinemia and high glycaemia reported in diseases and syndromes characterized by high insulinemia and glycaemia**

| Bacteria names | Results of this study | Diseases and syndromes characterized by high insulinemia and glycaemia |
|---|---|---|
| **Significantly different bacteria between individuals with low versus high insulin levels** | | |
| Phascolarctobacterium faecium | ↑ | ↑ Hypertension patient with and without type II diabetes[9] |
| Eubacterium eligens | ↓ | ↓ Type I diabetes[1], Gestational diabetes[2], Metabolic syndrome in HIV patients[3] |
| Bacteroides caccae | ↑ | ↑ Type II diabetes[4] |
| Bacteroides fragilis | ↑ | ↓ Type I diabetes in children[1], ↑ Type I diabetes in children[5], ↑ Obesity in children[6] |
| **Significantly different bacteria between individuals with low versus high glycaemia levels** | | |
| Intestinibacter bartlettii | ↓ | ↓ Overweight and obese children[10] |
| Bacteroides ovatus | ↓ | ↑ Type I diabetes[1], ↑ Type I diabetes in children[5], ↓ Obesity in children[7] |
| Fusicatenibacter saccharivorans | ↓ | - |
| Clostridium sp CAG 242 | ↑ | - |
| Firmicutes bacterium CAG 94 | ↑ | - |
| Streptococcus oralis | ↑ | ↑ Type I diabetes[8] |


1. Giongo A, Gano KA, Crabb DB, et al. Toward defining the autoimmune microbiome for type 1 diabetes. ISME J 2011;5(1):82–91.
2. Ma S, You Y, Huang L, et al. Alterations in Gut Microbiota of Gestational Diabetes Patients During the First Trimester of Pregnancy. Frontiers in Cellular and Infection Microbiology 2020;10(58):1–19.
3. Villanueva-Millán MJ, Pérez-Matute P, Recio-Fernández E, Lezana Rosales J-M, Oteo J-A. Characterization of gut microbiota composition in HIV-infected patients with metabolic syndrome. J Physiol Biochem 2019;75(3):299–309.
4. Zhong H, Ren H, Lu Y, et al. Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics. EBioMedicine 2019;47373–383.
5. de Goffau MC, Luopajärvi K, Knip M, et al. Fecal Microbiota Composition Differs Between Children With β-Cell Autoimmunity and Those Without. Diabetes 2013;62(4):1238–1244.
6. Ignacio A, Fernandes MR, Rodrigues VAA, et al. Correlation between body mass index and faecal microbiota from children. Clinical Microbiology and Infection 2016;22(3):258.e1-258.e8.
7. Maya-Lucas O, Murugesan S, Nirmalkar K, et al. The gut microbiome of Mexican children affected by obesity. Anaerobe 2019;5511–23.
8. Vatanen T, Franzosa EA, Schwager R, et al. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. Nature 2018;562(7728):589–594.
9. Ding H, Xu Y, Cheng Y, et al. Gut microbiome profile of Chinese hypertension patients with and without type 2 diabetes mellitus. BMC Microbiol 2023;23(1):254.
10. Murga-Garrido SM, Ulloa-Pérez EJ, Díaz-Benítez CE, et al. Virulence Factors of the Gut Microbiome Are Associated with BMI and Metabolic Blood Parameters in Children with Obesity. Microbiology Spectrum 2023;11(2):e03382-22.